



4-1-2021

## COPYRIGHTING COPYWRONGS: AN EMPIRICAL ANALYSIS OF ERRORS WITH AUTOMATED DMCA TAKEDOWN NOTICES

Seng, Daniel

Follow this and additional works at: <https://digitalcommons.law.scu.edu/chtlj>



Part of the [Intellectual Property Law Commons](#), and the [Science and Technology Law Commons](#)

### Recommended Citation

Seng, Daniel, *COPYRIGHTING COPYWRONGS: AN EMPIRICAL ANALYSIS OF ERRORS WITH AUTOMATED DMCA TAKEDOWN NOTICES*, 37 SANTA CLARA HIGH TECH. L.J. 119 (2021).

Available at: <https://digitalcommons.law.scu.edu/chtlj/vol37/iss2/1>

This Article is brought to you for free and open access by the Journals at Santa Clara Law Digital Commons. It has been accepted for inclusion in Santa Clara High Technology Law Journal by an authorized editor of Santa Clara Law Digital Commons. For more information, please contact [sculawlibrarian@gmail.com](mailto:sculawlibrarian@gmail.com).

## **COPYRIGHTING COPYWRONGS: AN EMPIRICAL ANALYSIS OF ERRORS WITH AUTOMATED DMCA TAKEDOWN NOTICES**

*By Daniel Seng<sup>1</sup>*

*Under the Digital Millennium Copyright Act (DMCA), reporters issuing takedown notices are required to identify the infringed work and the infringing material and provide their contact information (functional formalities), attest to the accuracy of such information and their authority to act on behalf of the copyright owner, and sign the notices (non-functional formalities). Online service providers will evaluate such notices for compliance with these DMCA formalities before acting on them. This paper seeks to answer questions about the quality of takedown notices, especially those generated by automated systems, which are increasingly being used by copyright owners to detect instances of online infringement and issue takedown notices on their behalf. After parsing three million takedown notices and more than eighty million takedown complaints served on Google between 2011 and 2015, this paper analyzes each notice for errors. This paper finds that almost all notices comply with the non-functional formalities. However, at least 5.5% of all takedown notices between*

---

<sup>1</sup> Daniel Seng, Associate Professor, Faculty of Law, National University of Singapore. I would like to thank Mark Lemley, Deborah Hensler, Phil Malone and Maria Jose Cordero for their useful feedback and suggestions for an earlier draft of this paper, which was entitled “Trust But Verify.” I would also like to record my deepest thanks to Shaun Lim for his invaluable help in revising this paper and updating the statistics with the porting of the Chilling Effects database onto the Lumen platform. The original material for this paper based on the Chilling Effects database was first presented at the 15th Year Retrospective of the Digital Millennium Copyright Act, Santa Clara University on March 15, 2013. It has since been updated with the dataset from the Lumen database. I also wish to thank Adam Holland from the Berkman Klein Centre for Internet & Society, Harvard University, for his kindness in enabling me unprecedented access to the Lumen database for purposes of this research. Funding support for the research that made this paper possible is provided by the National University of Singapore through the Singapore Ministry of Education Academic Research Fund Tier 1.

*2011 and 2015 fail to comply with the functional formalities in that they are missing copyright work descriptions. In addition, at least 9.8% of the takedown notices exhibit have empty takedown requests, misidentify the infringing site or provide inactive URIs as takedown requests. To ensure that the takedown system remains fast, efficient and error-free, this paper proposes to strengthen the attestation requirements of notices, to require reporters to validate all submitted takedown complaints and requests, and to subject recalcitrant reporters to the “slow lane” of a two-tier system for processing takedown notices. This methodology reflects the use of accountability metrics in the design of automated systems and suggests a verifiable response to address concerns pertaining to the use of systems that supplant human decision making.*

## CONTENTS

INTRODUCTION.....	122
I. STUDY METHODOLOGY.....	131
A. <i>The Datasets</i> .....	131
B. <i>Limitations of the Dataset</i> .....	135
II. COMPLIANCE WITH DMCA FORMALITIES.....	136
A. <i>The Two Classes of Notice Formalities under Section 512(c)(3)(A)</i> .....	136
B. <i>“Substantial Compliance” and “Technical Errors”</i> .....	139
C. <i>Errors in Non-functional Formalities</i> .....	146
D. <i>Errors in Functional Formalities</i> .....	151
E. <i>Summary</i> .....	164
III. SUBSTANTIVE ERRORS IN TAKEDOWN NOTICES.....	166
A. <i>Notices that Misidentify the Copyright Owner</i> .....	168
B. <i>The Megaupload Test</i> .....	171
IV. PROPOSALS FOR REFORM.....	183
CONCLUSION.....	190

## INTRODUCTION

When the Council of the European Union approved the Directive on Copyright in the Digital Single Market on 17 April 2019,<sup>2</sup> it was supposed to represent the latest updates to European copyright laws for the new digital age.<sup>3</sup> Instead, the application of the Copyright Directive is now in doubt: Poland has filed a legal challenge with the European Court of Justice and argued that its takedown-and-stay-down notice rule—Article 17 of the Copyright Directive<sup>4</sup>—undermined the “essence of the right to freedom of expression and information and [did] not comply with the requirement that limitations imposed on that right be proportional and necessary.”<sup>5</sup> The action by Poland reflects much of the controversy and intense lobbying activity in Europe that preceded the controversial passage of the Copyright Directive.<sup>6</sup> Indeed, it mirrors the online blackout in January 2012 that stemmed from concerns over how the Stop Online Piracy Act (SOPA) and the PROTECT IP Act (PIPA) introduced in the U.S. Congress would threaten free speech and innovation,<sup>7</sup> concerns similar to those raised

---

<sup>2</sup> Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, 2019 O.J. (L 130) 1, 1 [hereinafter *Copyright Directive*].

<sup>3</sup> *Id.* recital 83.

<sup>4</sup> *Id.* art. 17 (extending the existing takedown mechanism to a “takedown and staydown mechanism.”); *id.* recital 66. (“Additionally, such online content-sharing service providers should also be liable if they fail to demonstrate that they have made their best efforts to prevent the future uploading of specific unauthorised works, based on relevant and necessary information provided by rightholders for that purpose.”).

<sup>5</sup> Case C-401/19, Republic of Poland v. Council, 2019 O.J. (C 270).

<sup>6</sup> See, e.g., *EU Copyright Reform – What's so Controversial*, ABA (May 2019),

<https://www.americanbar.org/news/abanews/publications/youraba/2019/may-2019/eu-copyright-reform--whats-so-controversial/>.

<sup>7</sup> The Stop Online Piracy Act (SOPA) was a U.S. bill introduced in the House of Representatives with the goal of expanding the legal remedies available to content providers against Internet intermediaries to combat online copyright infringement. This included enabling the U.S. Attorney General to seek a court order to require “a service provider (to) take technically feasible and reasonable measures designed to prevent access by its subscribers located within the United States to the foreign infringing site.” H.R. 3261, 112th Cong.

concerning Article 17 of the Copyright Directive.<sup>8</sup> Bowing to these concerns, the U.S. Congress shelved SOPA and PIPA,<sup>9</sup> and retained the status quo in U.S. copyright law that is the Digital Millennium Copyright Act (DMCA).<sup>10</sup> Conversely, by passing Article 17 of the Copyright Directive, subject to the challenge from Poland, the Europeans succeeded in changing their existing copyright laws on intermediary liability that were largely based on the DMCA.<sup>11</sup> What was the underlying motivation for this unprecedented legislative reform of the DMCA?

In a nutshell, the DMCA, which was enacted in 1998, established the notice and takedown mechanism as part of an overall scheme to protect online service providers from fiscal liability for copyright infringement from their provision of services.<sup>12</sup> Under this notice and takedown mechanism, copyright owners will report instances of copyright infringement to the online service providers via takedown notices.<sup>13</sup> And service providers, who are under no general

---

§ 102(c)(2)(A)(i) (2011). [Herein after *Stop Online Piracy Act*]. The Senate version of SOPA was the Protect IP Act (Preventing Real Online Threats to Economic Creativity and Theft of Intellectual Property Act or PIPA). S. 968, 112th Cong. (2011). Concerns were raised that provisions such as these in SOPA and PIPA would enable law enforcement to block access to entire domains and lead to Internet censorship. This led to a coordinated effort by numerous Internet intermediaries to blackout their services on Jan. 18, 2012, in protest of these bills. See e.g. Amy Goodman, *The Sopa Blackout Protest Makes History*, THE GUARDIAN (Jan 18, 2012), <https://www.theguardian.com/commentisfree/cifamerica/2012/jan/18/sopa-blackout-protest-makes-history>.

<sup>8</sup> Case C-401/19, Republic of Poland v. Council, 2019 O.J. (C 270), <http://curia.europa.eu/juris/document/document.jsf?text=&docid=216823&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=19321452>.

<sup>9</sup> *Stop Online Piracy Act*, *supra* note 7. Plans to draft their alternative, the Online Protection & Enforcement of Digital Trade Act (OPEN), were also shelved in the U.S.

<sup>10</sup> *Id.*

<sup>11</sup> See, e.g., Miguel Peguera, *The DMCA Safe Harbors and Their European Counterparts: A Comparative Analysis of Some Common Problems*, 32 COLUM. J.L. & ARTS 481 (2009).

<sup>12</sup> See Senate Report 105-190: The Digital Millennium Copyright Act of 1998, at 20.

<sup>13</sup> See *id.* at 45.

obligation to monitor their services,<sup>14</sup> are to respond by taking down the allegedly infringing content referred to in the takedown notice.<sup>15</sup> Copyright owners have taken the stand that the online service providers are not doing enough under the takedown mechanism to fight copyright infringement.<sup>16</sup> Some owners have described the takedown mechanism as being ineffective by contending that for every takedown notice that they send to disable an infringing resource, multiple copies of that resource which the online service provider has taken down will surface—the so-called “whack-a-mole” problem.<sup>17</sup> Indeed, using the number of notices issued as an indicator of the scale of the problem of copyright infringement, copyright owners have exponentially increased the number of takedown notices they have issued across the years, from 125 thousand in 2011 to 1.31 million in 2015, which represents a year-on-year increase of 79.8% in notices *each year*.<sup>18</sup> Interestingly, in 2012, copyright owners ramped up their DMCA takedown notices by an astounding 88% *after* SOPA and PIPA failed.<sup>19</sup>

---

<sup>14</sup> 17 U.S.C. § 512(m)(1); *Copyright Directive*, *supra* note 2, art. 8; Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market 2000 O.J. (L 178) 1, 13 [hereinafter *E-Commerce Directive*].

<sup>15</sup> In this paper, the term “notice” is used to refer to the legal document addressed by the complainant to the receiving organization such as an online service provider or an individual. The term used in 17 U.S.C. § 512(c) is “notification.” Google’s Copyright Transparency Report rather confusingly refers to these notices as “requests,” whereas Twitter’s Transparency Report (correctly) refers to them as “notices.” *Google Search Removals Due to Copyright Infringement FAQs*, GOOGLE (2020), [http://www.google.com/transparencyreport/removals/copyright/faq/#what\\_is\\_a\\_copyright\\_removal\\_request](http://www.google.com/transparencyreport/removals/copyright/faq/#what_is_a_copyright_removal_request).

<sup>16</sup> See Brad Buckles, *Some Clear Facts About Google’s “Transparency” Report*, RECORDING INDUSTRY ASSOCIATION OF AMERICA (May 30, 2012), <https://www.benton.org/headlines/some-clear-facts-about-googles-transparency-report>.

<sup>17</sup> Section 512 Study: Notice and Request for Public Comment, 80 Fed. Reg. 81862 (Dec. 31, 2015), <https://www.copyright.gov/fedreg/2015/80fr81862.pdf> [hereinafter *Section 512 Study*].

<sup>18</sup> See *infra* Table 1.

<sup>19</sup> Daniel Seng, *The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices*, 18 VA. J.L. & TECH. 369, 390 (2013) [hereinafter

Using the number of takedown requests<sup>20</sup> as a proxy for enforcement action on the other hand shows that since 2011, there is an extraordinary year-on-year increase of 249% in requests *each year*.<sup>21</sup> If each takedown request represents a targeted unlicensed online resource, this means that the number of online resources targeted for takedowns has more than *doubled* each year. And on the basis that there are around 50 billion individually indexed web pages in the world in 2015,<sup>22</sup> assuming that there are no duplicate takedown requests, this means that in 2015, 565 million or at least 1.13% of web pages worldwide are alleged to contain unlicensed online resources.<sup>23</sup> This is an extraordinary statistic that represents the sheer scale and intensity of the enforcement action on the Internet.

---

*The State of the Discordant Union*]; see also Mike Masnick, *Funny How Copyright Holders Only Ramped Up Google DMCA Takedowns After SOPA Failed*, TECHDIRT (Dec. 13, 2012), <https://www.techdirt.com/articles/20121212/22445321369/funny-how-copyright-holders-only-ramped-up-google-dmca-takedowns-after-sopa-failed.shtml>.

<sup>20</sup> This paper uses the term “requests” to refer to the “information reasonably sufficient to permit the service provider to locate the material” in 17 U.S.C. § 512(c)(3)(A)(iii). Google’s Transparency Report somewhat confusingly also refers to these as “URLs requested to be removed” and “search results specified in requests.” See *Google Search Removals Due to Copyright Infringement FAQs*, *supra* note 15. This paper also uses the term “complaints” to refer to each complaint of infringement of copyright works for the same copyright owner or licensee, submitted in a single notice. For a further discussion, see Seng, *supra* note 19, at 401.

<sup>21</sup> See *infra* Table 1.

<sup>22</sup> *The Size of the World Wide Web (The Internet)*, WORLDWIDEBESIZE, (Dec. 19, 2020), <https://web.archive.org/web/20151219004022/https://www.worldwidewebsize.com/>.

<sup>23</sup> See *infra* Table 1.



Year	All Notices	Google Notices	All Complaints	Google Complaints	All Requests	Google Requests
2011	125,281	120,043	213,961	212,394	3,802,418	3,800,845
2012	531,414	527,329	2,388,048	2,375,073	57,201,313	57,188,338
2013	565,162	556,317	21,991,936	21,939,895	226,160,982	226,108,937
2014	865,446	845,587	29,127,087	29,105,686	317,133,198	317,110,917
2015	1,311,666	1,247,523	34,286,247	34,238,388	565,247,833	565,150,528

*Table 1: Notices, Complaints and Takedown Requests between 2011 and 2015 as reported to the Lumen database ( $N_2= 3,398,969$ )<sup>24</sup>*

On the other hand, it can be argued that the scale of these actions has achieved the objective of disabling access to infringing resources. Google has reported that notwithstanding the huge increases in notices and takedown requests received, takedown rates have been consistently high—at 97.5% from 2011–2012,<sup>25</sup> and above 98% in 2015.<sup>26</sup> Similarly, Microsoft reports that for the second half of 2018, the takedown rate for its Bing search engine is even higher, at 99.77%.<sup>27</sup> Outside of the realm of search engines, social media websites, like Twitter, also have consistently high takedown rates—at 61.8% in the second half of 2013 and at 74.0% in the second half of 2015.<sup>28</sup> In fact, the passage of Article 17 of the Copyright Directive and its takedown-and-stay-down mechanism are premised on the efficacy of existing takedown notices—with the added obligation on service providers to ensure that the materials disabled via the takedown notices

---

<sup>24</sup> Results obtained by way of a MySQL COUNT of second dataset (*see infra* text at note 54) of all notices, complaints and requests, and all notices, complaints and requests where recipient name is Google Inc.

<sup>25</sup> *How Google Fights Piracy*, GOOGLE, 14 (2013), <https://docs.google.com/file/d/0BwxyRPFduTN2dVfQYml5UENUeUE/edit> [hereinafter *How Google Fights Piracy*]; *see also* Fred von Lohmann, *Transparency for Copyright Removals in Search*, GOOGLE (May 24, 2012), <http://googleblog.blogspot.ca/2012/05/transparency-for-copyright-removals-in.html#!/2012/05/transparency-for-copyright-removals-in.html>.

<sup>26</sup> *How Google Fights Piracy*, GOOGLE, 38 (July 13, 2016), <https://drive.google.com/file/d/0BwxyRPFduTN2TmPqGajJ6TnRLaDA/view>.

<sup>27</sup> *Content Removal Requests Report*, MICROSOFT, <https://www.microsoft.com/en-us/corporate-responsibility/content-removal-requests-report>.

<sup>28</sup> *DMCA Takedown Notices*, TWITTER (Aug. 19, 2020), <https://transparency.twitter.com/en/reports/copyright-notice.html#2015-jul-dec>.

remain inaccessible.<sup>29</sup> All these suggest that contrary to the attempts made to reform the DMCA, the current takedown system of notices and requests is accurate, reliable and clearly effective.

But is it? The study most quoted on takedown notices is the 2006 empirical study by Jennifer Urban and Laura Quilter.<sup>30</sup> In reviewing 876 notices submitted to the Chilling Effects project through August 2005,<sup>31</sup> the authors found that at least one-third (37%) of the notices had substantive legal flaws relating to the underlying copyright claims or had significant noncompliance with the DMCA formalities for issuing takedown notices.<sup>32</sup> In a subsequent expanded study covering notices over a six-month period in 2013,<sup>33</sup> Urban *et al.* found that 4.2% of takedown requests were fundamentally flawed for mismatches in copyrighted and infringing content, and 28% of takedown requests were found to contain characteristics, such as statutory non-compliance and fair use, that raise clear questions about

---

<sup>29</sup> See *Copyright Directive*, *supra* note 2, Recital 66 para. 4 (recognizing this as part of the “high industry standards of professional diligence” expected of online service providers); see also Axel Voss, *Europäisches Urheberrecht* (June 18, 2018),

<https://web.archive.org/web/20180705234103/https://www.axel-voss-europa.de/2018/06/18/stellungnahme-zur-reform-des-urheberrechts/> (contending, as rapporteur of the Copyright Directive, for the further strengthening of the takedown mechanism by replacing it with a “takedown-and-stay-down” mechanism by requiring online platforms to take more responsibility on the basis that existing takedown and filtering mechanisms have been working successfully on platforms such as YouTube). There is, of course, the critical question of whether the stay-down mechanism is feasible, which is separate from whether the takedown itself is accurate.

<sup>30</sup> 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) [hereinafter *Efficient Process*].

<sup>31</sup> *Id.* at 641.

<sup>32</sup> *Id.* at 667. Google seems to agree with this figure. See *Internet Service Provider Copyright Code of Practice – TCF Consultation Draft*, GOOGLE, 9 n.3 (Mar. 6, 2009),

<https://web.archive.org/web/20160209121517/http://www.tcf.org.nz/content/ebc0a1f5-6c04-48e5-9215-ef96d06898c0.cmr>.

<sup>33</sup> Jennifer M. Urban et al., *Notice and Takedown in Everyday Practice*, (UC Berkeley Pub. Law Research Paper No. 2755628), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2755628](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755628) [hereinafter *Notice and Takedown*].

their validity.<sup>34</sup> And in a recent debate about SOPA and OPEN, a service provider representative claimed, without citing further evidence, that 5 to 10% of DMCA takedown notices are fraudulent.<sup>35</sup>

In the same vein, since Google began publishing daily postings of its takedown details, there has been limited systematic scrutiny of Google's takedown notices processed by various academic and internet observers.<sup>36</sup> This has consistently generated anecdotal reports of "abusive" notices<sup>37</sup> and notices that (intentionally or unintentionally) erroneously targeted legitimate content sites.<sup>38</sup> There were also

---

<sup>34</sup> *Id.* at 11–12.

<sup>35</sup> Jon Brodtkin, *Fighting Internet Piracy: CES Takes on SOPA vs. OPEN Debate*, ARSTECHNICA (Jan. 10, 2012, 5:10 PM), <http://arstechnica.com/tech-policy/2012/01/fighting-internet-piracy-ces-takes-on-sopa-vs-open-debate/>.

<sup>36</sup> See, e.g., *Research*, LUMEN, <https://lumendatabase.org/pages/research> (last visited Feb. 2, 2021).

<sup>37</sup> Abusive notices include those issued to seek the removal of unflattering information publicly posted about companies and individuals. See, e.g., *Unsafe Harbors: Abusive DMCA Subpoenas and Takedown Demands*, THE ELECTRONIC FRONTIER FOUNDATION (Sept. 26, 2003), [http://www.wipo.int/export/sites/www/meetings/en/2005/wipo\\_iis/presentations/pdf/wipo\\_iis\\_05\\_cdoctorow.pdf](http://www.wipo.int/export/sites/www/meetings/en/2005/wipo_iis/presentations/pdf/wipo_iis_05_cdoctorow.pdf); *Takedown Hall of Shame*, ELECTRONIC FRONTIER FOUNDATION, <https://www.eff.org/takedowns>; Paul Alan Levy, *A Bogus DMCA Takedown from Apple*, CONSUMER LAW & POLICY BLOG (Nov. 21, 2013), <http://pubcit.typepad.com/clpblog/2013/11/a-bogus-dmca-takedown-from-apple.html>; Ernesto Van der Sar, "Trolls" Try to Censor TorrentFreak's Copyright Trolls Coverage, TORRENTFREAK (Nov. 13, 2016), <https://torrentfreak.com/trolls-try-to-censor-torrentfreaks-copyright-trolls-coverage-161113/>.

<sup>38</sup> See, e.g., Ernesto Van der Sar, *Movie Studios Ask Google to Censor Their Own Films, Facebook and Wikipedia*, TORRENTFREAK (Dec. 3, 2012), <https://torrentfreak.com/movie-studios-ask-google-to-censor-their-own-films-facebook-and-wikipedia-121203/> (reporting that an anti-piracy organization known as "Yes It Is – No Piracy!" ostensibly sought to takedown legitimate copies of films on behalf of the movie studios on Verizon, Amazon, iTunes and so on, and even remove film reviews published by The Guardian, The Independent, The Mirror and the Daily Mail); Ernesto Van der Sar, *HBO Wants Google to Censor ... HBO.com*, TORRENTFREAK (Feb. 3, 2013), <http://torrentfreak.com/hbo-wants-google-to-censor-hbo-com-130203/>; Leigh Beadon, *You're All the Weakest Link: Bad Law Permits Bad Takedowns, Which Google Handles Badly*, TECHDIRT (Apr. 22, 2013, 2:21 PM), <http://www.techdirt.com/articles/20130422/09303922801/youre-all-weakest-link-bad-law-permits-bad-takedowns-which-google-handles-badly.shtml>;

falsified notices<sup>39</sup> that forced the reporters, who had allegedly served such notices, to claim that their systems were compromised<sup>40</sup> or were sent by imposters.<sup>41</sup> While errors are to be expected of a takedown system that is largely automated and driven by computers, occasionally,<sup>42</sup> the errors may be one too many. In September 2013, Microsoft terminated its partnership with leakID, a reporter, for

---

Andy Maxwell, *Thanks for the Really Counter-Productive DMCA Complaints*, TORRENTFREAK (May 24, 2015), <http://torrentfreak.com/thanks-for-the-really-counter-productive-dmca-complaints-150524/>; Ernesto Van de Sar, *Premier League Asks Google to Take Down Facebook's Homepage*, TORRENTFREAK (Apr. 29, 2017), <https://torrentfreak.com/premier-league-asks-google-to-take-down-facebooks-homepage-170429/>; Ernesto Van der Sar, *Rightsholders Remove Google Results of Legal Search Engine 'JustWatch'*, TORRENTFREAK (Aug. 11, 2019), <https://torrentfreak.com/rightsholders-remove-google-results-of-legal-search-engine-justwatch/>; Andy Maxwell, *DMCA Takedowns Try to Delist Dozens of Adult Homepages from Google*, TORRENTFREAK (June 8, 2019), <https://torrentfreak.com/dmca-takedowns-try-to-delist-dozens-of-adult-homepages-from-google-190608/>.

<sup>39</sup> See, e.g., Eric Limer, *Prankster Uses Bogus DMCA Notices to Takedown Bieber's YouTube Channel* (Sept. 1, 2011, 4:40 PM), <https://www.themarysue.com/bogus-dmca-beiber-gaga/>; Jon Brodtkin, *How a Single DMCA Notice Took Down 1.45 Million Education Blogs*, ARS TECHNICA (Oct. 15, 2012, 3:00 PM), <http://arstechnica.com/information-technology/2012/10/how-a-single-dmca-notice-took-down-1-45-million-education-blogs/>; Ernesto Van der Sar, *Microsoft DMCA Notice 'Mistakenly' Targets BBC, Techcrunch, Wikipedia and U.S. Govt*, TORRENTFREAK (Oct. 7, 2012), <http://torrentfreak.com/microsofts-bogus-dmca-notices-censor-bbc-cnn-wikipedia-spotify-and-more-121007/>.

<sup>40</sup> See, e.g., Andy Maxwell, *Anti-Piracy Co. Blames Hack for Bogus DMCA's, But They're Just Sloppy*, TORRENTFREAK, (Mar. 7, 2012), <http://torrentfreak.com/anti-piracy-co-blames-hack-for-bogus-dmcas-but-theyre-just-sloppy-120307/>.

<sup>41</sup> See, e.g., Ernesto Van der Sar, *Scammers Hit Pirate Game Sites With 'Irreversible' Google Takedowns*, TORRENTFREAK (Nov. 30, 2018), <https://torrentfreak.com/scammers-hit-pirate-game-sites-with-irreversible-google-takedowns-181130/>; Ernesto Van der Sar, *'Netflix' Flags Netflix.com As a Pirate Site, Or Does It?* TORRENTFREAK (June 2, 2019), <https://torrentfreak.com/netflix-flags-netflix-com-as-a-pirate-site-or-does-it-190602/>; Ernesto Van der Sar, *Fake MPAA Asks Google to Remove Thousands of URLs, Including MPAA.org*, TORRENTFREAK (July 14, 2019), <https://torrentfreak.com/fake-mpaa-asks-google-to-remove-thousands-of-urls-including-mpaa-org-190714/>.

<sup>42</sup> See, e.g., Disney Enterprises, Inc. v. Hotfile Corp., No. 11-20427-CIV, 2013 WL 6336286 at \*15 (S.D. Fla. Sept. 20, 2013).

sending out inaccurate notices, including notices that targeted Microsoft's Wikipedia entry and even Microsoft's own website for piracy.<sup>43</sup> That is not to say that leakID (and Microsoft) were not previously warned. Although their identities were not publicly disclosed, it is believed<sup>44</sup> that leakID was one of two reporters who had been ejected by Google a year earlier from its Trusted Copyright Removal Program for Web Search (TCRP) for repeatedly sending inaccurate notices.<sup>45</sup> Should Microsoft have taken heed from leakID's eviction from the TCRP? Or should leakID itself have cleaned up its act after its loss of TCRP status?

So, what is the true state of the quality of takedown notices today? Do the reporting mechanisms of owners and reporters identify infringing materials and activities accurately? Surely that is a prerequisite to a takedown regime and is critical to the success of a takedown-and-stay-down regime.<sup>46</sup> In fact, if the quality of takedown notices is poor, the stay-down regime will reinforce the concerns of pundits that it will operate as a censorship system that will stymie the freedom of expression and information.<sup>47</sup> Thus, it is crucial to identify the types of errors made by owners and reporters when issuing and processing takedown notices and the rates of these errors. The explosion of takedown notices is because of the use of sophisticated

---

<sup>43</sup> Ernesto Van der Sar, *Microsoft Ditches Anti-Piracy Partner After Embarrassing DMCA Takedowns*, TORRENTFREAK (Sept. 27, 2013), <http://torrentfreak.com/microsoft-ditches-anti-piracy-partner-after-embarrassing-dmca-takedowns-130927/>.

<sup>44</sup> This is from a close examination of the contents of LeakID's issued takedown notices as stored in the Chilling Effects Lumen database. The indicative date of termination of LeakID's partnership is Aug. 23, 2012. *Cf.*, for instances, the following notices, *Software DMCA (Copyright) Complaint to Google*, LUMEN (Aug. 22, 2012), <https://www.lumendatabase.org/notices/1221009>, *with DMCA (Copyright) Complaint to Google*, LUMEN (Aug. 23, 2012), <https://www.lumendatabase.org/notices/1228118>.

<sup>45</sup> *How Google Fights Piracy*, *supra* note 25, at 17.

<sup>46</sup> *See* S. REP. NO. 105-190, at 46 (1998); H.R. REP. NO. 105-551, pt. 2, at 55 (1998) (referring to the need for the adequate identification information to "reasonably sufficient to permit the service provider to identify and locate the allegedly infringing material").

<sup>47</sup> *See, e.g.*, Daniel Nazer, *Copyright, The First Wave of Internet Censorship*, ELECTRONIC FRONTIER FOUNDATION (Jan. 18, 2018), <https://www EFF.org/deeplinks/2018/01/copyright-first-wave-internet-censorship>.

tools and automated means that use digital fingerprinting, hash values and keyword or metadata searches to identify unlicensed content that is being disseminated—the so-called “robo-takedowns.”<sup>48</sup> Can we trust these mechanisms that operate with little or no human intervention?<sup>49</sup>

This paper examines these questions in four parts. Part I explains the methodology used for extracting the notices addressed to Google for this case study. Part II analyzes the notices for their compliance with DMCA formalities (“formal errors”). Part III analyzes the notices for errors that go to the substance of the claims embedded in notices (“substantive errors”). And Part IV concludes with some proposals for legal reform to address the formal and substantive errors studied in this paper.

## I. STUDY METHODOLOGY

### A. *The Datasets*

The methodology used in this paper builds on the study methodology used in an earlier paper,<sup>50</sup> which accessed the Chilling Effects repository for takedown notices.<sup>51</sup> The first dataset was built by collating individual takedown notices submitted to the Chilling Effects repository. For this dataset, a cut-off date of December 31, 2012 was used. After parsing the notices, and filtering only for notices submitted to Google, this yielded slightly more than half-a-million ( $N_1=501,286$ ) fully parsed notices submitted to Google, comprising 56,991,045 takedown requests.<sup>52</sup>

After the Chilling Effects repository was superseded by the Lumen repository,<sup>53</sup> the second dataset was built by collating individual takedown notices submitted by all participating online

---

<sup>48</sup> Zoe Carpou, *Robots, Pirates, and the Rise of the Automated Takedown Regime: Using the DMCA to Right Piracy and Protect End-Users*, 39 COLUM. J. L. & ARTS 551 (2016).

<sup>49</sup> Brief of Amicus Curiae Electronic Frontier Foundation in Support of Hotfile’s Opposition to Summary Judgment on its Counterclaim at 19, *Disney Enterprises, Inc. v. Hotfile Corp.*, No. 11-20427-CIV, 2013 WL 6336286 (S.D. Fla. Sept. 20, 2013).

<sup>50</sup> Seng, *supra* note 19, at 378–83.

<sup>51</sup> The repository is of takedown notices between 2009 and 2012, referred to in this paper as the first dataset.

<sup>52</sup> Seng, *supra* note 19, at 383.

<sup>53</sup> *About Us*, LUMEN, <https://lumendatabase.org/pages/about> (last visited Feb. 2, 2021) [hereinafter *Lumen – About Us*].

service providers,<sup>54</sup> including Google, Twitter<sup>55</sup> and Microsoft,<sup>56</sup> to the Lumen repository.<sup>57</sup> The second dataset used cut-off dates of between January 1, 2011 and December 31, 2015. Like the first dataset, a census<sup>58</sup> of all copyright-related DMCA notices submitted to Google was conducted.<sup>59</sup> This yielded more than three million ( $N_2=3,398,969$ ) notices (including 3,296,799 notices submitted to Google), comprising 88.0 million takedown complaints and 1.17 *billion* takedown requests.<sup>60</sup> The second dataset is primarily used to analyze the notices, complaints and requests for formal and substantive errors.

---

<sup>54</sup> LUMEN, <https://lumendatabase.org/>.

<sup>55</sup> *Lumen – About Us*, *supra* note 53 (listing, among others, Google Inc. and Twitter, Inc. as contributors to the Lumen repository).

<sup>56</sup> *See, e.g., Search*, LUMEN, [https://lumendatabase.org/notices/search?utf8=%E2%9C%93&recipient\\_name=Bing&sort\\_by=](https://lumendatabase.org/notices/search?utf8=%E2%9C%93&recipient_name=Bing&sort_by=) (showing, since Feb. 3, 2020, 30,225 results of notices submitted to Microsoft Bing) (last visited Feb. 3, 2021).

<sup>57</sup> LUMEN, <https://lumendatabase.org/>.

<sup>58</sup> Conducting a census will also enable a longitudinal analysis to be made of various issues investigated.

<sup>59</sup> These would be those tagged using the Lumen meta data as relating to “dmca” disputes.

<sup>60</sup> More exactly, 88,007,279 takedown complaints and 1,169,545,744 requests were detected in the second dataset. Results obtained by way of a MySQL COUNT of second dataset of all notices, complaints and requests.

Year	Average Complaints/ Notice	SD Complaints/ Notice	Max Complaints/ Notice	Average Requests/ Notice	SD Requests/ Notice	Max Requests/ Notice
<b>All Notices (N<sub>2</sub>=3,299,265)</b>						
2011	1.71	8.36	997	30.35	152.27	9,967
2012	4.49	32.97	1,943	107.64	665.48	25,050
2013	38.91	138.85	1,998	400.17	1,411.95	23,147
2014	33.66	126.05	2,000	366.44	1,442.74	23,426
2015	26.14	103.55	1,513	430.94	1,628.81	21,687
<b>Google notices only (N=3,296,799)</b>						
2011	1.77	8.51	997	31.66	155.42	9,967
2012	4.50	33.09	1,943	108.45	667.99	25,050
2013	39.44	139.88	1,998	406.44	1,422.25	23,147
2014	34.42	127.42	2,000	375.02	1,458.48	23,426
2015	27.45	106.01	1,513	453.02	1,667.17	21,687
<b>Non-Google notices only</b>						
2011	0.31	2.85	90	0.31	2.85	90
2012	3.29	6.38	40	3.29	6.38	40
2013	5.96	8.79	108	5.96	8.79	108
2014	1.08	3.63	20	1.12	4.86	271
2015	0.75	0.44	1	1.52	6.42	362

Table 2: Average, Standard Deviation and Maximum Number of Complaints and Requests per Notice between 2011 and 2015, for All Notices, Google Notices and Non-Google Notices (N<sub>2</sub>=3,299,265)<sup>61</sup>

While the first and second datasets overlap for the years 2011 and 2012, the second dataset shows a significant change in the way takedown requests are reported to online service providers. To begin, unlike the notices reviewed in the 2006 study,<sup>62</sup> there has been a significant increase in the use of complaints as “sub-notices” within each notice since 2011.<sup>63</sup> Averaging 1.71 complaints per notice in 2011, this figure rose to 4.49 complaints per notice in 2012 and by 2015, it had reached 26.14 complaints per notice.<sup>64</sup> The number of

<sup>61</sup> Results obtained by using MySQL to conduct on the second dataset a SUM of COUNT of complaints and requests for each notice for each year and conducting basic analysis to compute the average, standard deviation and maximum statistics.

<sup>62</sup> See Urban & Quilter, *supra* note 30, at 658–63. The description of notices in *Efficient Process* implies that most of the notices then (pre-2006) examined contained a takedown request for either a single work or a closely-related set of works such as a portfolio; this is empirically observed to no longer hold true with the present dataset.

<sup>63</sup> See *supra* Table 2.

<sup>64</sup> See *supra* Table 2.



takedown requests per notice has also risen from an average of 30.35 requests per notice in 2011 to 430.94 requests per notice in 2015.<sup>65</sup>

The rise in the use of complaints within notices and the increase in takedown requests per notice appear to be prompted in no small part by Google's introduction of the TCRP program sometime in March 2011, which permitted TCRP members to make bulk submissions of takedown notices and requests directly to Google.<sup>66</sup> The ratio of complaints to notices increased even further with the lifting of the cap on takedown requests by Google in April 2013.<sup>67</sup> This had the effect of dramatically increasing the maximum number of requests per notice. For instance, while the maximum number of requests in a notice was 9,967 in 2011, since 2012, the number has not fallen below 20,000.<sup>68</sup> As the table above shows,<sup>69</sup> all these records of complaints and requests per notice are set only by Google notices: non-Google notices have witnessed much smaller averages and maximum complaints and requests per notice, often smaller by several orders of magnitude than Google notices.

The ability to pack so many requests into a single takedown notice has both encouraged and in turn facilitated the use by reporters of automated tools to detect infringing resources and submit large numbers of these takedown requests in an automated manner to online service providers for quick action.<sup>70</sup> The voluminous quantities of takedown requests coupled with the speed of takedowns has therefore increased the urgency of independently validating the veracity of these takedown notices, complaints and requests.

---

<sup>65</sup> While the average number of complaints and requests per notice appear reasonable, the standard deviations are high, suggesting a large amount of variance in the number of complaints and requests per notice over the whole dataset. An examination of the distribution reveals that a small number of reporters are responsible for sending a disproportionate majority of notices containing large numbers of complaints or requests, but there are many one-off or infrequent reporters who send small notices dealing with isolated instances of infringement. *See* Seng, *supra* note 19, at 397.

<sup>66</sup> Seng, *supra* note 19, at 414–16.

<sup>67</sup> *Id.*

<sup>68</sup> *See supra* Table 2.

<sup>69</sup> *See supra* Table 2.

<sup>70</sup> *See* Perfect 10, Inc. v. Giganews, Inc., No. 11-CV-07098-AB (SHx), 2015 WL 1746406, at \*10 (C.D. Cal. Mar. 6, 2015).

### B. *Limitations of the Dataset*

The limitations of the first dataset were explored in an earlier paper.<sup>71</sup> In particular, because this study focuses exclusively on Google's takedown notices, the observations made here may not relate easily to other online service providers. Also, both datasets do not include takedown notices received by Google as part of its YouTube Content ID system.<sup>72</sup>

An earlier iteration of this paper made exclusive use of the first dataset, which was derived from Lumen's predecessor, Chilling Effects.<sup>73</sup> When it was sought to repeat the same analysis using the second dataset, which was derived from Lumen, it was discovered that there are differences in how Lumen and Chilling Effects record and present takedown notices. As a result, certain analyses possible on the Chilling Effects dataset are not possible on the Lumen dataset. For instance, while Chilling Effects recorded data related to statutory formalities, such as DMCA attestations as separate fields, Lumen does not. These differences will be pointed out as and when they occur in a form germane to the results.

In addition, in the process of analyzing these notices, data formatting inconsistencies and truncation of the contents of a small number of Google notices (fewer than a thousand) in the first dataset were detected.<sup>74</sup> To the extent that the DMCA attestations were appended to the end of notices, the corruption could explain the many instances of attestation errors.<sup>75</sup> However, to the extent that the notices are missing the requisite attestations, this paper reports them as they are.

Finally, notwithstanding the efforts of the Chilling Effects and Lumen repositories to redact the names and identities of individual

---

<sup>71</sup> Seng, *supra* note 19, at 383–88.

<sup>72</sup> The Content ID system works because participating content providers have submitted eighty million digital fingerprints of their audio and video works to YouTube, which are matched against every uploaded video. See *Mission Report: Towards More Effectiveness of Copyright Law on Online Content Sharing Platforms: Overview of Content Recognition Tools and Possible Ways Forward*, FRENCH MINISTRY OF CULTURE 16 (Jan. 29, 2020), <https://www.culture.gouv.fr/content/download/265045/file/Mission%20Report%20Content%20Recognition%20Tools%20ENG%20V.pdf?inLanguage=fr-e-FR>.

<sup>73</sup> See *Lumen – About Us*, *supra* note 53.

<sup>74</sup> Seng, *supra* note 19, at 388.

<sup>75</sup> See *infra* note 114 and accompanying text.

parties from the takedown notices,<sup>76</sup> the parsers developed for this study managed to extract from notices some individual names and identities that were not redacted.<sup>77</sup> Reference will be made to some of these individual names in this paper to illustrate the points about mistakes in notices, but in this paper, these individuals' names have been scrambled to preserve their identities and protect their privacy.

## II. COMPLIANCE WITH DMCA FORMALITIES

### A. *The Two Classes of Notice Formalities under Section 512(c)(3)(A)*

Section 512 provides that an online service provider shall be exempted from monetary relief if, among others, it responds expeditiously to remove, or disable access to, allegedly infringing material or activity,<sup>78</sup> upon an “effective notification” of claimed infringement.<sup>79</sup> To be an effective notification, it must be “a [single]<sup>80</sup> written communication”<sup>81</sup> provided to the “designated agent”<sup>82</sup> of a service provider that includes “substantially” the following six formalities:

- (i) a physical or electronic signature of a person authorized to act on behalf of the owner of an exclusive right that is allegedly infringed (“notice signature”);<sup>83</sup>
- (ii) the identification of the copyrighted work claimed to have been infringed, or a representative list of such works, if multiple copyrighted works at a single site are targeted in the notice (“description of the copyrighted work”);<sup>84</sup>

---

<sup>76</sup> See *Lumen Tools for Researchers*, LUMEN (2017), <https://www.lumendatabase.org/pages/researchers#database>.

<sup>77</sup> See *infra* note 217 and accompanying text.

<sup>78</sup> 17 U.S.C. §§ 512(b)(2)(E), (c)(1)(C), (d)(3).

<sup>79</sup> See S. REP. NO. 105-190, at 45–46 (1998).

<sup>80</sup> *Perfect 10, Inc. v. Google, Inc.*, No. CV 04-9484 AHM (SHx)SHX, 2010 WL 9479059, at \*11 (C.D. Cal. July 26, 2010).

<sup>81</sup> *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102, 1113 (9th Cir. 2007) (noting that the notice has to be in the form of a single and not separate pieces of written communication).

<sup>82</sup> 17 U.S.C. § 512(c)(2).

<sup>83</sup> 17 U.S.C. § 512(c)(3)(A)(i).

<sup>84</sup> 17 U.S.C. § 512(c)(3)(A)(ii).

- (iii) the identification of the material claimed to be infringing or to be the subject of infringing activity and that is to be removed or access to which is to be disabled, and information reasonably sufficient to permit the service provider to locate the material (“takedown request”<sup>85</sup>);<sup>86</sup>
- (iv) information reasonably sufficient to permit the service provider to contact the complaining party, such as an address, telephone number, and if available, the complaining party's electronic mail address (“reporter’s contact information”);<sup>87</sup>
- (v) a statement that the complainant “has a good faith belief that use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law” (“statement of good faith belief”);<sup>88</sup> and
- (vi) a statement that “the information in the notification is accurate, and under penalty of perjury, that the complaining party is authorized to act on behalf of the owner of an exclusive right that is allegedly infringed” (“statement of accuracy and authorization”).<sup>89</sup>

Exact compliance with the formal requirements is not required: the test is one of *substantial* compliance.<sup>90</sup> A notice that fails to comply substantially with the formalities need not be acted on by the recipient service provider as it is not “effective.”<sup>91</sup> And an ineffective notice cannot be considered to form the corpus of the service provider's actual knowledge or “red flag knowledge” of infringing activity.<sup>92</sup> It will also not trigger the service provider's obligations with respect to its “repeat

---

<sup>85</sup> See *supra* note 20 and accompanying text.

<sup>86</sup> 17 U.S.C. § 512(c)(3)(A)(iii).

<sup>87</sup> 17 U.S.C. § 512(c)(3)(A)(iv).

<sup>88</sup> 17 U.S.C. § 512(c)(3)(A)(v).

<sup>89</sup> 17 U.S.C. § 512(c)(3)(A)(vi).

<sup>90</sup> 17 U.S.C. § 512(c)(3)(A).

<sup>91</sup> *Id.*

<sup>92</sup> 17 U.S.C. § 512(c)(3)(B)(i); see S. REP. NO. 105-190 at 45 (1998) (explaining that neither actual knowledge nor awareness of a red flag may be imputed to a service provider based on information from a copyright owner, or its agents, that does not comply with the notification provisions set out in § 512(c)(3). In such a circumstance, the service provider's indemnity from fiscal damages is intact.); see also S. REP. NO. 105-190 at 46–47 (1998).

infringer policy.”<sup>93</sup> However a savings clause in the DMCA provides that an “ineffective” notice that otherwise substantially complies with (ii), (iii) and (iv) may be considered in evaluating a service provider’s knowledge of infringing activity unless it “promptly attempts to contact [the complainant] or takes other reasonable steps to assist in the receipt of notification that substantially complies” with all the aforesaid formal requirements.<sup>94</sup> Presumably, the service provider need not act on such an “ineffective” notice after it has contacted the complainant unless the complainant subsequently rectifies it.<sup>95</sup>

This means that while every notice has to “substantially” meet *all six* formalities,<sup>96</sup> in terms of the *consequences* of non-compliance, the formalities may be divided into two broad classes: the “functional”<sup>97</sup> or “one-shot only” formalities—description of the copyrighted work, the takedown request and the reporter’s contact information (items (ii), (iii) and (iv))—and the “non-functional” or “second-shot possible” formalities—the notice signature, the statement of good faith belief, and the statement of accuracy and authorization (items (i), (v) and (vi)).<sup>98</sup> (The two statements are also referred to in this paper as “the attestations.”). Rewording the explanation above, the difference between the two classes of formalities is that if there is no substantial compliance with the functional formalities, the notice fails *in limine*, but where a notice fails to substantially comply with the non-functional formalities, the service provider is obliged to contact the complainant and give her a second chance to remedy these defects.<sup>99</sup>

---

<sup>93</sup> Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1111 (9th Cir. 2007).

<sup>94</sup> 17 U.S.C. § 512(c)(3)(B)(ii); see S. REP. NO. 105-190, at 46–47 (1998).

<sup>95</sup> See S. REP. NO. 105-190, at 47 (1998) (explaining that if the service provider subsequently receives a substantially complainant notice, it would be required to act expeditiously on it); cf. Urban & Quilter, *supra* note 3033, at 674 (suggesting that an online service provider is not exempt from responding to a notice with errors in the non-functional formalities).

<sup>96</sup> See Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1112 (9th Cir. 2007) (“substantial compliance means substantial compliance with all of § 512(c)(3)’s clauses, not just some of them.”).

<sup>97</sup> See S. REP. NO. 105-190, at 47 (1998); see also H.R. REP. NO. 105-551, pt. 2, at 56 (1998) (using the term “functional requirements” to describe those formalities where sufficient information must be supplied to ensure that the notification and take down procedures may operate).

<sup>98</sup> See 17 U.S.C. § 512(c)(3)(B)(ii).

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

B. “Substantial Compliance” and “Technical Errors”

In *Perfect 10, Inc. v. CCBill LLC*, the court held that a notice that is *completely missing* one or more of the required elements cannot be said to substantially comply with the requirements of section 512(c)(3).<sup>100</sup> The court considered this to be beyond a “technical error.”<sup>101</sup> So what is the standard of “substantial compliance” which notices have to meet with respect to each of the formalities, assuming that it is not missing any of these formalities? The DMCA does not elucidate.<sup>102</sup>

The Senate and House Reports note that “substantial compliance” includes compliance in which “technical errors” such as misspelling a name, supplying outdated area codes for phone numbers if the numbers are accompanied by accurate addresses, or supplying outdated names (of the prior designated agent) if accompanied by valid e-mail addresses for the successor agent are disregarded.<sup>103</sup> Aside from these examples, which relate only to the reporter’s contact information (item (iv)), the Reports offer no further guidance. But the DMCA caselaw offers additional insights.

In *Rosen v. Hosting Services Inc.*, the court rejected notices that erroneously identified the copyrighted work claimed to have been infringed (item (ii)).<sup>104</sup> In *Perfect 10, Inc. v. Google, Inc.*, the court likewise rejected notices that broadly referred to more than 15,000 images appearing on the plaintiff’s website as lacking the identification of the copyrighted work claimed to have been infringed.<sup>105</sup>

The court also rejected notices with requests that specify top-level URLs and truncated URLs (item (iii)).<sup>106</sup> The court opined that an

---

<sup>100</sup> *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102, 1112 (9th Cir. 2007) (“The statute thus signals that substantial compliance means substantial compliance with all of § 512(c)(3)’s clauses, not just some of them.”).

<sup>101</sup> *Id.*

<sup>102</sup> *Recording Indus. Ass’n of Am., Inc. v. Verizon Internet Servs., Inc.*, 351 F.3d 1229, 1236 (D.C. Cir. 2003).

<sup>103</sup> S. REP. NO. 105-190, at 47 (1998); H.R. REP. NO. 105-551, pt. 2, at 56 (1998).

<sup>104</sup> *Rosen v. Hosting Servs., Inc.*, 771 F. Supp. 2d 1219, 1222 (C.D. Cal. 2010).

<sup>105</sup> *Perfect 10, Inc. v. Google, Inc.*, No. CV 04-9484 AHM (SHx), 2010 WL 9479059, at \*10–12 (C.D. Cal. July 26, 2010).

<sup>106</sup> *Id.* at 10, 12, 14 (rejecting the Group C notices issued against Google’s Search and Blogger services, because the URLs were incomplete and were instead truncated by ellipses).

incomplete URL is not a specific link that affords identification of the infringing material.<sup>107</sup> Likewise, in *Perfect 10, Inc. v. Giganews, Inc.*, the court rejected notices that direct the recipient service provider to conduct a particular search for the allegedly infringing Usenet message. It held that search results rather than URIs (which encompass URLs and Usenet message identifiers)<sup>108</sup> would not enable the service provider to unambiguously identify and locate the Usenet message in question.<sup>109</sup>

From the aforesaid DMCA cases, it is possible to construe a more general rule for determining when an incomplete (as opposed to a missing) formality will cease to be substantial compliance. Case law has generally interpreted “substantial” compliance with reference to “something less than a strict and literal compliance.”<sup>110</sup> Merriam-Webster's Collegiate Dictionary defines “substantial” as “being largely but not wholly that which is specified.”<sup>111</sup> But while some deviation is permitted, the requisite level of “compliance” is one that has to meet

---

<sup>107</sup> *Id.* at 13, 12 n.9.

<sup>108</sup> A URI is an Internet address, and as an engineering unit, it unambiguously resolves to a specific location where online resources can be found. The Uniform Resource Indicator (URI) is a string of characters used to identify a name or an Internet resource. A URI can be a Uniform Resource Locator (URL or web address) or a Uniform Resource Name (URN), such as a Usenet message-ID. A URI therefore includes a URL. For an explanation of the difference between a URI and a URL, see MICHAEL MEALLING & RAY DENENBERG, REPORT FROM THE JOINT W3C/IETF URI PLANNING INTEREST GROUP: UNIFORM RESOURCE IDENTIFIERS (URIs), URLs, AND UNIFORM RESOURCE NAMES (URNs): CLARIFICATIONS AND RECOMMENDATIONS, (The Internet Society 3 (2002), available at <http://tools.ietf.org/html/rfc3305>). In this study, for the most part, the two terms can be used interchangeably, since the resources referred to here are primarily web pages and resources. For a longer discussion, see Seng, *supra* note 19, at 401–02.

<sup>109</sup> *Perfect 10, Inc. v. Giganews, Inc.*, 993 F. Supp. 2d 1192, 1200-1201 (C.D.Cal. 2014) (drawing the analogy between a Message-ID which is the only unique identifier to locate Usenet messages and the URL).

<sup>110</sup> *Wells Benz, Inc. v. United States*, 333 F.2d 89, 92 (9th Cir.1964) (explaining that substantial compliance “does imply something less than a strict and literal compliance with the contract provisions but fundamentally it means that the deviation is unintentional and so minor or trivial as not ‘substantially to defeat the object which the parties intend to accomplish.’”).

<sup>111</sup> *Definition of Substantial*, MERRIAM WEBSTER (2020), <https://www.merriam-webster.com/thesaurus/substantial>.

the “essential statutory purpose”<sup>112</sup> of the formalities or procedural provisions. The following quotation is often cited to explain the requirement of “substantial compliance”:

“Substantial compliance” with a statute means *actual compliance in respect to the substance essential to every reasonable objective of the statute*. It means that a court should determine whether the statute has been followed sufficiently so as to carry out the intent for which it was adopted. Substantial compliance with a statute is not shown unless it is made to appear that the purpose of the statute is shown to have been served. What constitutes substantial compliance with a statute is a matter *depending on the facts of each particular case*. (emphasis added).<sup>113</sup>

What is “substantial compliance” with the formalities, therefore, is a question of identifying the extent of deviation from “strict and literal compliance” and the statutory purpose behind each formality deviated from. Here, it is vital to understand the context in which takedown notices operate today. As the first paper demonstrated,<sup>114</sup> gone are the days where copyright owners and reporters submit a few notices each day, which are then reviewed manually for errors by service providers. With higher volumes of takedown requests from “robo-requests” and faster turnarounds expected of many service providers, it is submitted that the types of errors tolerated will become smaller. This will explain why the Senate Report uses the term “technical” error to describe those errors which are accepted as being in “substantial compliance”—misspelled names, outdated area codes for phone numbers accompanied by accurate addresses, or outdated names accompanied by valid e-mail addresses.<sup>115</sup> These are errors which service providers can afford the indulgence of oversight because at the time the U.S. Congress enacted the DMCA, “less than 5% of the world’s population

---

<sup>112</sup> American Air Filter Co, v. Commissioner of Internal Revenue, 81 T.C. 709, 719 (1983) (“substantial compliance with regulatory requirements may suffice when such requirements are procedural and when the essential statutory purposes have been fulfilled.”).

<sup>113</sup> Wagner v. Truesdell, 574 N.W.2d 627, 629 (S.D. 1998).

<sup>114</sup> Seng, *supra* note 19, at 389 (noting that Google received less than 1,000 takedown notices in 2009, which was more than 10 years after its incorporation).

<sup>115</sup> See S. REP. No. 105-190, at 47 (1998).



used the internet,”<sup>116</sup> and it did not “anticipate the online world as we now know it—where, each day, users post hundreds of millions of photos, videos and other items, and service providers receive over a million notices of alleged infringement.”<sup>117</sup>

If so, in the context of today’s world of automated takedown notices, the term “technical errors” has to receive a narrower definition. It is submitted that “technical errors” have to be errors that are “endogenously detectable” and “endogenously remediable.” An error is “detectable” if the existence of the error can be detected. It is “endogenously detectable” if it is erroneous or ambiguous on its face; unlike an error which is “exogenously detectable” if reference is made to external resources to ascertain that the information is in error. An error must be first “detectable” before it can be “remedied,” though detecting it is no assurance that it can. It is “endogenously remediable” if it can be corrected based on the existing information provided on the notice, and it is “exogenously remediable” if it can only be corrected with reference to external resources.

A close examination will show that the “technical error” examples provided in the Senate Report<sup>118</sup> are instances of errors relating to the reporter’s contact information that are, using the terms defined above, “endogenously detectable” and “endogenously remediable.” For instance, where the correct telephone area code can be inferred from an accurate address that is supplied, or where the name of the prior reporting agent can be disregarded because the new reporter’s e-mail address is supplied,<sup>119</sup> these will count as “technical errors.”

But the varying instances of “technical compliance” must always take into account the statutory object of each of the formalities. This is so particularly as regards the takedown request for “identification of the material claimed to be infringing.”<sup>120</sup> As noted above, supplying the abbreviated, truncated, or misspelled URI is invariably going to lead to a non-compliant notice<sup>121</sup> because of the critical need to be accurate, precise, and unambiguous with the

---

<sup>116</sup> *Section 512 Study, supra* note 17, at 81862.

<sup>117</sup> *Id.*

<sup>118</sup> *See* S. REP. No. 105-190, at 47 (1998).

<sup>119</sup> *Id.*

<sup>120</sup> 17 U.S.C. § 512(c)(3)(A)(iii).

<sup>121</sup> *Perfect 10, Inc. v. Google, Inc.*, No. CV 04-9484 AHM (SHx), 2010 WL 9479059 at \*10, \*12, \*14 (C.D. Cal. July 26, 2010).

takedown request.<sup>122</sup> There is no way to unambiguously “guess” the correct URI from a misspelt or erroneous URI. To have it otherwise could lead to the removal of non-infringing material belonging to an innocent user, and the termination of speech protected under the First Amendment.<sup>123</sup>

In relation to this requirement, the U.S. Copyright Office in its Section 512 Study in 2015 made the following observation:

Since the passage of the DMCA, courts have been called upon to address the elements required for an “effective”—*i.e.*, valid—take down notice. Looking to section 512's requirement to provide “information reasonably sufficient to permit the service provider to locate the material,” courts have generally required a high degree of specificity, such as the particular link, or uniform resource locator (“URL”), where the infringing material is found. Likewise, service providers often request that the specific URL for each allegedly infringing use be included in a notice. Such a requirement can be *burdensome* in the case of a notice that references a large number of infringements at multiple locations throughout the same site. Additionally, copyright owners question whether this level of specificity is in conflict with the statute's express language allowing complaining parties to submit a “representative list” of works alleged to be infringed “at a single online site.”<sup>124</sup>

This observation surely cannot be correct. Firstly, the Copyright Office erroneously cites as support for this observation, 17 U.S.C. § 512(c)(3)(A)(ii), which relates to describing the copyrighted work,

---

<sup>122</sup> It is for this reason that the lower court in *Viacom Intern. Inc. v. YouTube, Inc.* noted that “a copy or description of the allegedly infringing material and the URL address of the web page location which is alleged to contain the infringing material” is an example of sufficient information. *See Viacom Intern. Inc. v. Youtube, Inc.*, 940 F. Supp. 2d 110, 115 (S.D.N.Y. 2013).

<sup>123</sup> *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102, 1112 (9th Cir. 2007) (“Accusations of alleged infringement have drastic consequences: A user could have content removed or may have his access terminated entirely. If the content infringes, justice has been done. But if it does not, speech protected under the First Amendment could be removed.”).

<sup>124</sup> *Section 512 Study*, *supra* note 17, at 81865.

rather than the infringing work.<sup>125</sup> The Copyright Office seems to have obscured the distinction between *infringed* works (i.e., copyrighted works) with *infringing* works, contrary to the distinction maintained in 17 U.S.C. § 512(c)(3)(A)(ii) ("copyrighted work claimed to have been infringed") and § 512(c)(3)(A)(iii) ("material . . . claimed to be infringing"). Secondly, if multiple locations of a site harbor allegedly infringing resources, the onus should remain on the copyright owners to identify these locations and thus these resources. The DMCA is clear in that it does not require the service provider to conduct affirmative monitoring.<sup>126</sup> The specificity of the URI operates as notification by the copyright owner to the online service provider by imputing to the service provider actual knowledge of specific and identifiable infringement, which in turn triggers the service provider's obligation to effect an expeditious removal or disablement of access to that resource.<sup>127</sup> The requirement to provide a URI is thus consistent with the entire takedown mechanism in the DMCA.

Turning to the item (ii) or "description of the copyrighted work" claimed to have been infringed, as noted above, the DMCA allows a representative list of copyrighted works to be submitted "if multiple copyrighted works at a single online site are covered by a single notification."<sup>128</sup> Presumably, there will be "substantial compliance" if this list omits some copyrighted works that are claimed to have been infringed by the single online site targeted by the takedown request. However, this exception has been narrowly construed to foreclose on a complaining party making a broad and indiscriminating reference to *all* its copyrighted works as failing to provide the requisite identification of the copyrighted work claimed to have been infringed.<sup>129</sup> The rationale for this must be that the concept of a representative list is only tenable when the Internet, as conceived when the DMCA was enacted in 1998, was small and the number of

---

<sup>125</sup> *Id.* at 81865 n.41.

<sup>126</sup> 17 U.S.C. § 512(m); *see* S. REP. No. 105-190, at 48-49 (1998); H.R. REP. No. 105-551, pt. 2, at 57-58 (1998).

<sup>127</sup> *Viacom Int'l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 30 (2d Cir. 2012) ("Thus, the nature of the removal obligation itself contemplates knowledge or awareness of specific infringing material, because expeditious removal is possible only if the service provider knows with particularity which items to remove.").

<sup>128</sup> 17 U.S.C. § 512(c)(3)(A)(ii).

<sup>129</sup> *Perfect 10, Inc. v. Google, Inc.*, No. CV -04-984 AHM (SHx), 2010 WL 9479059, at \*10-12 (C.D. Cal. July 26, 2010).

instances of copyright infringement were manageable manually.<sup>130</sup> This is no longer the case, as can be seen by the numbers of takedown notices sent by complaining parties using automated processes<sup>131</sup> and processed by online service providers through the use of scalable systems to address the large volumes of notices.<sup>132</sup> If the object behind the “description of the copyrighted work” formality is to impute to the online service provider the requisite knowledge of the specific instance of infringement, there must also be the requisite specificity in identifying the work being infringed.<sup>133</sup>

It is also well known that some infringing works are deliberately misspelt to avoid or minimize detection by copyright owners and their agents.<sup>134</sup> Prominent examples include “Micro\$oft” and “Windoze” as intentional misspellings of Microsoft and Windows.<sup>135</sup> Perhaps some misspellings may be tolerated to some degree, if any ambiguities can be endogenously resolved (e.g., there is no doubt that “Micro\$oft Windoze” refers to the Windows operating system from Microsoft Corporation, particularly where Microsoft is noted as the copyright owner in the notice). But to relax the requirements of substantial compliance any further may be to require the service provider to play mind reading games to figure out what the

---

<sup>130</sup> *Section 512 Study*, *supra* note 17, at 81862 (“At that time, less than 5% of the world’s population used the internet . . .”).

<sup>131</sup> *Id.* at 81864.

<sup>132</sup> *Id.* at 81865.

<sup>133</sup> *See* *Viacom Int’l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 30 (2nd Cir. 2012); *see also* *Rosen v. Hosting Servs., Inc.*, 771 F. Supp. 2d 1219, 1222 (2010).

<sup>134</sup> *See, e.g., Napster Faced with Big List, Trick Names*, ABC NEWS (Jan. 6, 2006), <https://abcnews.go.com/Entertainment/story?id=108389&page=1>; *see also* *A&M Records, Inc. v. Napster, Inc.*, 2001 WL 227083, at \*1 (N.D. Cal. Mar. 5, 2001), *aff’d*, 284 F.3d 1091 (9th Cir. 2002) (“All parties shall use reasonable measures in identifying variations of the filename(s), or of the spelling of the titles or artists’ names, of the works identified by plaintiffs. If it is reasonable to believe that a file available on the Napster system is a variation of a particular work or file identified by plaintiffs, all parties have an obligation to ascertain the actual identity (title and artist name) of the work and to take appropriate action within the context of this Order.”).

<sup>135</sup> THE NEW HACKER’S DICTIONARY, VERSION 4.3.3 (2002), <https://www.landley.net/history/mirror/jargon.html#The%20Jargon%20Lexicon>.

copyright owner or reporter intended to take down.<sup>136</sup> Consequently, this will encroach on the sacrosanct principle that a service provider is under no obligation to monitor “its service or affirmatively seek[] facts indicating infringing activity.”<sup>137</sup> “The DMCA notification procedures place the burden of policing copyright infringement—identifying the potentially infringing material and adequately documenting infringement—squarely on the owners of the copyright.”<sup>138</sup> Given that information identifying the infringed material may form the basis for imputing actual or constructive knowledge of infringing activity upon the service provider, it is submitted that compliance with this formality has to be strict, and “substantial compliance” has to be narrow. Support for this strict interpretation of the formality for description of the copyrighted work can be found in the case of *Rosen v. Hosting Services, Inc.*, where the court held that a takedown notice which misidentifies the allegedly infringed material is defective and as a matter of law, the recipient service provider could not “be charged with having the requisite knowledge to be contributorily liable.”<sup>139</sup>

### C. *Errors in Non-functional Formalities*

So, what is the state of non-functional formalities compliance of takedown notices based on the test of substantial compliance? Prior to the use of automated systems leading to the present explosion in takedown notices and web forms, the 2006 Urban and Quilter study<sup>140</sup> found such “statutory flaws” in “one out of every eleven notices,” but limited its analysis to the functional formalities, and did not include an analysis of the non-functional formalities.<sup>141</sup> However, a notice can now comprise thousands of complaints, and each complaint can comprise tens of thousands of takedown requests.<sup>142</sup> Similarly, the use of web forms has prevailed over all other forms of takedown notices.<sup>143</sup>

---

<sup>136</sup> A mismatch between the name of the allegedly infringing work and the copyright work would *prima facie* not appear to be an infringement from the name alone; a hypothetical reporter might have to provide both the correct name of the work, its misspelled (or even disguised) name, alongside further substantiation as to why the allegedly infringing work is infringing.

<sup>137</sup> 17 U.S.C. § 512(m)(1).

<sup>138</sup> *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102, 1113 (2007).

<sup>139</sup> *Rosen v. Hosting Servs., Inc.*, 771 F. Supp. 2d at 1222.

<sup>140</sup> *See* Urban & Quilter, *supra* note 30, at 621.

<sup>141</sup> *Id.* at 674.

<sup>142</sup> *See* Seng, *supra* note 19, at 434.

<sup>143</sup> *Id.* at 398–400.

Have these changes affected the rate of compliance with the DMCA formalities rules?

This study starts first with the “non-functional” or “second-shot possible” formalities: the notice signature, the statement of good faith belief and the statement of accuracy and authorization. That these are “non-functional” formalities does not make them any less important. As the court in *CCBill* explained:

The DMCA requires a complainant to declare, under penalty of perjury, that he is authorized to represent the copyright holder, and that he has a good-faith belief that the use is infringing. This requirement is not superfluous. Accusations of alleged infringement have drastic consequences: A user could have content removed or may have his access terminated entirely. If the content infringes, justice has been done. But if it does not, speech protected under the First Amendment could be removed. We therefore do not require a service provider to start potentially invasive proceedings if the complainant is unwilling to state under penalty of perjury that he is an authorized representative of the copyright owner, and that he has a good-faith belief that the material is unlicensed.<sup>144</sup>

Because “substantial compliance” of these non-functional formalities may take a myriad of forms, for the purpose of this empirical study, to avoid interpretational issues as to what constitutes “substantial compliance,” especially disputes as to whether there is “substantial compliance” in the particular context of the contents of a notice and for that formality in issue, an “error” with a formality is narrowly defined as one where the formality examined is *completely missing* from the notice. So defined, a missing formality cannot be in compliance, let alone be in substantial compliance. Likewise, a notice with a missing formality cannot be said to be in compliance with the DMCA.<sup>145</sup>

The analysis below deals with data from the first dataset, since it was discovered that only the Chilling Effects database retains information on the non-functional formalities as earlier explained.<sup>146</sup>

---

<sup>144</sup> Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1112 (9<sup>th</sup> Cir. 2007).

<sup>145</sup> 17 U.S.C. § 512(c)(3)(A).

<sup>146</sup> See *supra* note 73 and accompanying text.

Using the broad definition of an “electronic signature,”<sup>147</sup> where a notice does not have an electronic symbol (“signed,” “signature” or its variants) or a process (such as an appended signature date, or appended name of the copyright owner or reporter) associated with it, it will be flagged as in error for missing an electronic signature (“Missing Notice Signature”).

For the attestations, as is the case in *CCBill*, where the notice does not contain a statement that the complainant “has a good faith belief that use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law” (“Missing Statement of Good Faith Belief”), or a statement that “the information in the notification is accurate” (“Missing Statement of Accuracy”), or a statement that “under penalty of perjury . . . the complaining party is authorized to act on behalf of the owner of an exclusive right that is allegedly infringed” (“Missing Statement of Authorization”) (and variations of these statements) (there being two separate requirements in this formality), it will be flagged as missing that formality.<sup>148</sup>

All these flagged notices are then individually verified (thus eliminating the problem of false positives).<sup>149</sup> This analysis will therefore pick up a *subset* of the notices with formalities which are *not* in substantial compliance: the results from the analysis will represent the *lower* bound of notices that are in error. This yields the following table:

---

<sup>147</sup> 15 U.S.C. § 7006(5) incorporating the Electronic Signatures in Global and National Commerce Act 2000, 114 Stat. 464 (“electronic signature” defined as “an electronic sound, symbol, or process, attached to or logically associated with a contract or other record and executed or adopted by a person with the intent to sign the record”); Uniform Electronic Transactions Act 1999, § 2(8) (“electronic signature” defined as “an electronic sound, symbol, or process attached to or logically associated with a record and executed or adopted by a person with the intent to sign the record.”).

<sup>148</sup> *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102, 1112 (9th Cir. 2007).

<sup>149</sup> As there were fewer than 400 of such erroneous notices, it was possible to conduct a manual verification of all the flagged notices. Note that the verification process will not be able to determine false negative instances, e.g. if there is a notice which does not have a signature, which the parsing process does *not* flag as missing a signature. So understood, these figures represent the *minimum* percentage of notices which are missing the formalities (the lower bound); actual figures may be *higher*.

	2008	2009	2010	2011	2012
<b>All Form Notices</b>	6	2,457	7,279	57,490	435,869
<b>Missing Notice Signature – item (i)</b>	0 (0.000%)	15 (0.595%)	16 (0.215%)	151 (0.260%)	153 (0.035%)
<b>Missing Statement of Good Faith Belief – item (v)</b>	0 (0.000%)	5 (0.198%)	1 (0.013%)	56 (0.097%)	69 (0.016%)
<b>Missing Statement of Accuracy – item (vi)</b>	0 (0.000%)	4 (0.159%)	2 (0.027%)	50 (0.086%)	69 (0.016%)
<b>Missing Statement of Authorization – item (vi)</b>	0 (0.000%)	4 (0.159%)	3 (0.040%)	60 (0.104%)	70 (0.016%)

Table 3: Notices with Non-functional Formalities Errors (Signatures and Attestations – items (i), (v) and (vi)) between 2008 and 2012 ( $N_1=501,286$ )<sup>150</sup>

This study reveals quite a different picture since the 2006 Urban and Quilter study. Setting aside the figures for 2008 (at only six notices in the dataset, there were too few of them to draw any definitive conclusions), the figures show that the rate of errors for these non-functional formalities has decreased and is now extremely low, even as the number of form-based takedown notices received by Google has increased exponentially.<sup>151</sup> Starting with 2009, out of 2,457 notices, only 0.60% of Google’s form-based notices were missing signatures.<sup>152</sup> In 2010, out of 7,279 Google notices, only 0.22% of them were missing signatures.<sup>153</sup> By 2012, out of 435,869 Google notices, only 0.035% of them did not have signatures.<sup>154</sup> So, notwithstanding an almost eight-fold increase in notices from 2011 to 2012, the numbers of notices with missing signatures showed an almost eight-fold *decrease* during the same period.<sup>155</sup>

The rates of errors with the attestations show a similar reduction. In 2009, 0.020% of all notices have a Missing Statement of

---

<sup>150</sup> Results obtained by using MySQL to conduct on the first dataset a COUNT of notices by their form, signatures, statements of good faith, statements of accuracy and statements of authorization, for each of the years.

<sup>151</sup> See *supra* Table 3.

<sup>152</sup> See *supra* Table 3.

<sup>153</sup> See *supra* Table 3.

<sup>154</sup> See *supra* Table 3.

<sup>155</sup> See *supra* Table 3.



Good Faith, a Missing Statement of Accuracy or a Missing Statement of Authorization attestation.<sup>156</sup> By 2012, only 0.016% of these notices have a Missing Statement of Good Faith, a Missing Statement of Accuracy or a Missing Statement of Authorization attestation.<sup>157</sup> So, notwithstanding a 177-fold *increase* in notices received between 2009 and 2012, there has been a 10-fold *decrease* in notices with missing attestations during the same period.<sup>158</sup>

That so many form-based notices would be free of “non-functional” formalities errors should *not* come as a surprise.<sup>159</sup> Google’s form submission page requires the complainant to check boxes acknowledging the accuracy of the notices, affirming her good faith and her authority to act before the submission can be processed.<sup>160</sup> In fact, the converse is surprising: that given the prevalence of web form-based notices, there should be *any* notices at all with formal errors. A manual review of notices with Missing Statement of Good Faith, Missing Statement of Accuracy or Missing Statement of Authorization attestations showed that many of these notices had truncated contents, especially at the section which marked the complainant’s sworn statements. If this is due to some form of transmission or conversion error between Google and Chilling Effects, the actual numbers of notices with “non-functional” formalities could be *lower*.

---

<sup>156</sup> See *supra* Table 3.

<sup>157</sup> See *supra* Table 3.

<sup>158</sup> See *supra* Table 3.

<sup>159</sup> These results are validated by Urban’s updated study, which, in sampling 1,827 notices from 6 months of takedown notices submitted to the Chilling Effects database in 2013 (amounting to a total of 108 million takedown requests), found only “a handful of requests” which did not entirely conform to the statutory requirements, and which found that all requests in their sample were submitted through online forms. See Urban et al., *supra* note 33 at 93.

<sup>160</sup> Google’s Web Search takedown form requires complainants to “check to confirm” the aforesaid attestations. A failure to check the attestation fields will produce an error message “Required field must not be blank” and the takedown notice cannot be submitted. See *Removing Content from Google*, GOOGLE: LEGAL HELP

<https://support.google.com/legal/troubleshooter/1114905?hl=en#ts=9814647%2C1115655%2C9814950%2C1115789%2C1117010%2C1697925> (select Create Request to view form).

*D. Errors in Functional Formalities*

What about notices with errors as regards the description of the copyrighted work, the takedown request and the reporter's contact information—the functional formalities errors? As noted above, notices which have functional formalities errors render them substantially non-compliant and they cannot be subsequently rectified by the reporter.

To analyze the notices with functional formalities errors, notices in the second dataset will be parsed for missing entries for identification of the copyright work, identification of the infringed material and identification of the reporter. Because a notice may comprise several complaints of infringement of copyright works for the same copyright owner or licensee, where a complaint is missing any description of the copyright work, this study will flag the complaint (and the notice in its entirety)<sup>161</sup> as non-compliant, because, in such a case, it is not possible to relate the supplied takedown requests in the complaint to the infringed copyright work. An example of such a notice (with information as to the description of the copyrighted work and the original URL of the copyrighted work) is as follows:

---

<sup>161</sup> It may be argued that the failure to include a description for the copyright work only invalidates that complaint, but not the entire notice. However, the attestation, under penalty of perjury, that “the information in the notification is accurate” relates to the entire *notice*, and not just to the complaint. 17 U.S.C. § 512(c)(3)(A)(ii). In addition, the concept of a “complaint” is not recognized in the DMCA: it is a tool of convenience created by online service providers, predominantly, Google Inc., to facilitate the submission of takedown notices for reporters representing copyright owners for a large group of works. *See infra* text accompanying note 16.

**DMCA (Copyright) Complaint to Google**

<p><b>SENDER</b>  <b>leakid</b>  on behalf of <b>microsoft</b>  [Private]  ..., FR  Sent on June 15, 2012</p>	<p><b>RECIPIENT</b>  <b>Google, Inc.</b>  [Private]  ...</p>
---	--

OTHER ENTITIES: PRINCIPAL

**Re: Websearch Infringement Notification via Online Form Complaint**  
SENT VIA: ONLINE FORM

**NOTICE TYPE:** DMCA

**Copyright claim #1**

**KIND OF WORK:** Unspecified

**DESCRIPTION**

**ORIGINAL URLS:** No copyrighted URLs were submitted.

**ALLEGEDLY INFRINGING URLS:**

- 01. pastebin.com - 5 URLs
- 02. www.filescopy.com - 4 URLs
- 03. rapidsharegames.net - 4 URLs
- 04. rapidlibrary.com - 4 URLs
- 05. www.linktury.com - 3 URLs

*Figure 1: Example of Notice with Empty Copyright Work Description: Notice 130280 dated June 15, 2012, submitted by leakID to Google<sup>162</sup>*

Where the complaint does not have a takedown request or URI, it (and the notice in its entirety) will be flagged as missing information that identifies and enables the location of the infringing work.<sup>163</sup> The most typical example of such a notice is one that has a complaint without *any* takedown request relating to the work. The following notice (with missing information as to the “Allegedly Infringing URLs”) is an example:

<sup>162</sup> *Websearch Infringement Notification via Online Form Complaint*, LUMEN (June 15, 2012), <https://lumendatabase.org/notices/130280#>.

<sup>163</sup> See Seng, *supra* note 19, at 401–04.

DMCA (Copyright) Complaint to Google	
<b>SENDER</b> Harsh Patel [Private] ... US Sent on June 06, 2011	<b>RECIPIENT</b> Google, Inc. [AdSense] [Private] Mountain View, CA, 94043, US
OTHER ENTITIES: PRINCIPAL	
<b>Re: Infringement Notification via AdSense Complaint</b> SENT VIA: ONLINE FORM: FORM	
<b>NOTICE TYPE:</b>	DMCA
<b>Copyright claim #1</b>	
<b>KIND OF WORK:</b>	Unspecified
<b>DESCRIPTION</b>	The work is a set of instructions completely written by me. All images and text on the site are copyrighted. In the footer of each page (including the page of the copyrighted work) includes this statement "No part of this blog or its feed may be republished without the written consent of the author." The footer also includes a link titled Legal Info. The linked page ( <a href="http://ipodtoucher55.blogspot.com/2008/11/legal-notice.html">http://ipodtoucher55.blogspot.com/2008/11/legal-notice.html</a> )- <a href="http://www.google.com/url?sa=D&amp;q=http%3A%2F%2Fipodtoucher55.blogspot.com%2F2008%2F11%2Flegal-notice.html">http://www.google.com/url?sa=D&amp;q=http%3A%2F%2Fipodtoucher55.blogspot.com%2F2008%2F11%2Flegal-notice.html</a> )> includes more information on the copyright of the entire site.
<b>ORIGINAL URLS:</b>	01. www.google.com - 1 URL 02. ipodtoucher55.blogspot.com - 1 URL
<b>ALLEGEDLY INFRINGING URLS:</b>	No infringing URLs were submitted.
<a href="#">Click here</a> to request access and see full URLs.	
<b>JURISDICTIONS</b>	Unspecified
<b>Original Documents</b>	

Figure 2: Example of Notice with Complaint without Takedown Request: Notice 152065 dated Jun. 6, 2011, submitted by Harsh Patel to Google<sup>164</sup>

Bearing in mind the distinction noted above between notices and complaints, the following table sets out the statistics for all Google notices and their complaints that potentially exhibit functional errors:

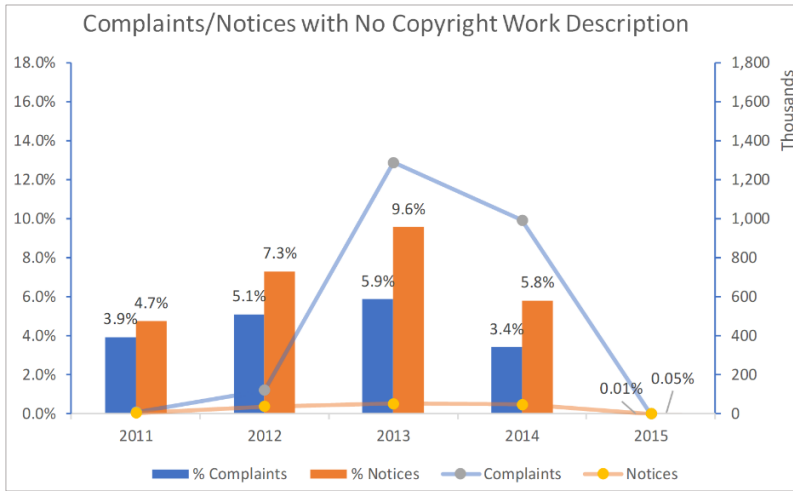
<sup>164</sup> Websearch Infringement Notification via Online Form Complaint, LUMEN (Jun. 6, 2011), <https://lumendatabase.org/notices/152065>.

Year	2011	2012	2013	2014	2015
<b>All Google Takedown</b>	212,394 complaints	2,375,073 complaints	21,939,895 complaints	29,105,686 complaints	34,238,388 complaints
<b>Notices: notices/complaints</b>	102,176 notices	512,143 notices	551,048 notices	822,670 notices	1,246,755 notices
<b>No Copyright Work Description - item (ii)</b>	8,386 complaints (3.95%)	121,053 complaints (5.10%)	1,288,178 complaints (5.87%)	992,280 complaints (3.41%)	2,036 complaints (0.006%)
(% of All Google)	4,851 notices (4.75%)	37,467 notices (7.32%)	52,720 notices (9.57%)	47,827 notices (5.81%)	593 notices (0.048%)
<b>As Above + No Copyright Work URLs - (% of All Google/% of No Copyright Work Description)</b>	5,941 complaints (2.80%/70.8%)	119,699 complaints (5.04%/98.9%)	1,287,968 complaints (5.87%/100%)	990,530 complaints (3.40%/99.8%)	198 complaints (0.001%/9.72%)
	3,761 notices (3.68%/77.5%)	36,119 notices (7.05%/96.4%)	52,513 notices (9.53%/99.6%)	47,487 notices (5.77%/99.3%)	198 notices (0.016%/33.4%)
<b>No Takedown Request in any complaint - item (iii)</b>	12,288 complaints (5.79%)	134,750 complaints (5.67%)	1,317,585 complaints (6.01%)	1,585,618 complaints (5.45%)	1,637,288 complaints (4.78%)
(% of All Google)	9,801 notices (9.59%)	43,454 notices (8.48%)	68,175 notices (12.4%)	93,564 notices (11.4%)	91,291 notices (7.32%)
<b>As Above + TCRP senders only</b>	3,250 complaints (1.53%/26.4%)	113,489 complaints (4.78%/84.2%)	1,297,429 complaints (5.91%/98.5%)	1,544,496 complaints (5.31%/97.4%)	1,604,323 complaints (4.69%/98.0%)
(% of All Google/% of No Takedown Request)	1,346 notices (1.32%/13.7%)	32,710 notices (6.39%/75.3%)	53,099 notices (9.64%/77.9%)	79,933 notices (9.72%/85.4%)	80,044 notices (6.42%/87.7%)

Table 4: Notices with Formal Errors (items (ii) and (iii)) between 2011 and 2015 ( $N_2=3,299,265$ )<sup>165</sup>

The numbers above can be better shown in the following graphs.

<sup>165</sup> Results obtained by using MySQL to conduct on the second dataset a COUNT of complaints for each notice for each year based on the various criteria – missing copyright work descriptions (with missing URLs to describe the copyright work) and missing requests (with senders who are TCRP participants).



*Figure 3: Plot of Percentage of Notices and Complaints with No Copyright Work Descriptions between 2011 and 2015<sup>166</sup>*

The statistics show that there are substantial numbers of non-compliant complaints that fail to supply a copyright work description. The percentage of non-compliant complaints rose from 3.95% (representing 4.75% of notices) in 2011 to 5.87% (representing 9.57% of notices) in 2013, before falling to 3.41% (representing 5.81% of notices) in 2014, and further to 0.006% (representing 0.048% of notices) in 2015.<sup>167</sup> Given that Google's current online forms for submitting takedown requests mandate that this information be

<sup>166</sup> See *supra* Table 4.

<sup>167</sup> See *supra* Figure 3.

supplied,<sup>168</sup> it is surprising that there are submitted notices which fail to comply with this formality.<sup>169</sup>

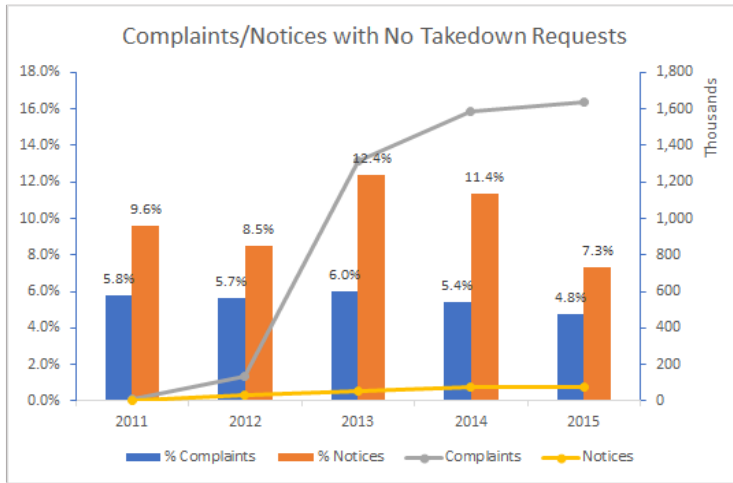


Figure 4: Plot of Percentage of Notices and Complaints with No Takedown Requests between 2011 and 2015<sup>170</sup>

Nor is the situation any much better in relation to information for identifying the allegedly infringing material. As Table 4:4 shows, the reporters did not fare very well here either. It is found that 9.59% of notices in 2011 have one or more complaints that do not have *any*

<sup>168</sup> Urban's updated study opined that webforms used by Google would increase the likelihood that the required formalities were observed. Urban et al., *supra* note 33, at 93. The introduction of tightened rules for submitting takedown notices might explain the sharp reduction in notices and complaints without copyright work descriptions in 2015. But as Urban's study only examined notices in 2013, the results above represent the first quantitative evidence of the positive effects of such submission rules that strictly enforce the DMCA formalities. However, it might also be contended that if the submission rules were tightened, there should be *no* instances of complaints or notices with no copyright work descriptions. The presence of a small number of such notices/complaints might be because these were submitted before the tightened submission rules were implemented in 2015.

<sup>169</sup> By cross-referencing the same notice as found on the first dataset (from the Chilling Effects repository) and the second dataset (from the Lumen repository), it can be shown that the same notices that do not have any copyright descriptions across both datasets and both repositories. Of course, one possible, but unlikely, explanation for this is that the contents of the same notices were corrupted when published onto both repositories. It is submitted that the better explanation is that these are real mistakes made by reporters.

<sup>170</sup> See *supra* Table 4.

valid URI information, rising to 12.37% in 2013 before falling to 7.32% in 2015.<sup>171</sup>

If a complaint is missing the description of the copyright work, it may be argued that the failure to include a description for the copyright work only invalidates that complaint and its underlying requests, but not the entire notice. On this view, the other complaints with complete descriptions and requests remain valid. The better view starts with the observation that only Google and Twitter give reporters the option to submit multiple complaints for each notice.<sup>172</sup> Reporters do submit single-complaint notices to Google and Twitter, and if they submit multiple complaints in one notice out of expediency, they must also take the burden of ensuring that all the complaints in that one notice are accurate. After all, all reporters have to support *each notice* with the attestation, under penalty of perjury, that “the information in the notification is accurate” (statement of accuracy).<sup>173</sup> Enabling a reporter to select *some* complaints and the attestation for the notice as support for a notification that is substantially compliant shades into the objection against enabling a reporter to put together a complaint notice from separately defective notices as this would unduly burden the service provider.<sup>174</sup> As the court in *Perfect 10, Inc. v. CCBill LLC* noted, “the text of § 512(c)(3) requires that the notice be ‘a written communication.’”<sup>175</sup> Thus, it is submitted that a complaint that is missing the copyright work description should be rendered non-compliant and so should its parent notice.

It is submitted that a complaint with no takedown request will taint not only that complaint, but the entire notice and all its associated complaints (which may have takedown requests), for the same reasons outlined above. A complaint without any request takes away the *raison d’être* for the notice, which is an “accusation[] of alleged infringement”

---

<sup>171</sup> See *supra* Table 4.

<sup>172</sup> Results obtained by using MySQL to conduct on the second dataset a COUNT of notices that have more than one complaint and isolating the recipient names for such notices.

<sup>173</sup> 17 U.S.C. § 512(c)(3)(A)(vi).

<sup>174</sup> *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102, 1112–13 (9th Cir. 2007).

<sup>175</sup> *Id.* at 1113 (observing that this requirement for the notice to be one single written communication instead of separate communications is not a mere technicality as otherwise, this would require the service provider to piece together the relevant information for each instance of claimed infringement, which would shift a substantial burden of policing and documenting the infringement away from the copyright owner to the service provider).



that the service provider is obliged to act on to “start potentially invasive proceedings.”<sup>176</sup> It bears emphasis that every notification has to be supported by the reporter’s statement of accuracy *and* a statement of his good faith belief that “use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law”<sup>177</sup> (the statement of good faith belief). A notice with *some* complaints that have *no* takedown requests will leave the service provider uncertain about the provenance of the reporter’s good faith belief about the entire notice, and unclear about the possible knowledge imputed to the provider from the rest of the complaints in the notice.<sup>178</sup>

There is some anecdotal evidence that these errors could be attributed to the unsophisticated reporter. For instance, with some types of takedown forms,<sup>179</sup> individual reporters may not know what a URI is.<sup>180</sup> Some supply only descriptive information in place of a URI.<sup>181</sup> Others are confused by the requirement to supply not only the URI of the “infringing material in the catalog” (the reference to “catalog” is totally misleading), but also the URI of “the infringing third party content that the blog is linking to” (this refers to third party material that is linked to the page complained of as above). This has led to no end of confusion for some individual reporters,<sup>182</sup> many of whom have

---

<sup>176</sup> *Id.* at 1112.

<sup>177</sup> 17 U.S.C. § 512(c)(3)(A)(v).

<sup>178</sup> 17 U.S.C. § 512(c)(3)(B)(i) (imputing to the service provider actual knowledge or awareness of facts or circumstances of infringing activity from a substantially compliant notification).

<sup>179</sup> Notably, the non-Google Search forms.

<sup>180</sup> *See, e.g., Blog DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (June 7, 2011) (on file with author) (stating that for the field that requires the reporter to enter the “Location (URL) of infringing material in the catalog,” the reporter entered, “It’s only on your server.”); *see also, e.g., Music DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Nov. 10, 2012) (on file with author) (stating that the reporter entered “google” when asked to supply the “URL of the allegedly infringing material in our search results”).

<sup>181</sup> *See, e.g., DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Oct. 29, 2012) (on file with author) (stating “daddyfunplace 'chubby in mini speedo' safe search off” when asked to supply the “URL of the allegedly infringing material in our search results”).

<sup>182</sup> *See, e.g., Music DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (May 4, 2010) (on file with author) (stating that for the field that requires the reporter to enter the “Location (URL) of infringing third party content that the blog is linking to,” the reporter entered, “I don’t know what in hell this means . . .”).

vented their frustrations in the notice itself.<sup>183</sup> This suggests that the unsophisticated users do require some instruction as to what constitutes a valid URI in a takedown request.

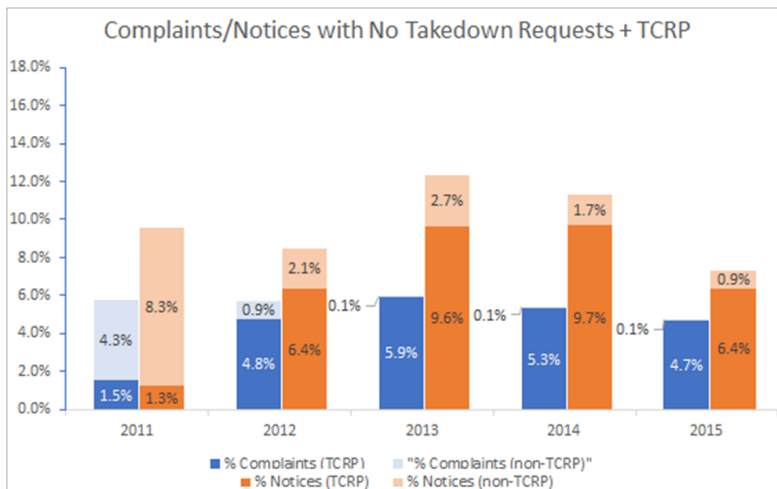
But the figures in Table 4 confirm that almost all of these notices with “no URI” complaints are issued by “trusted users”—sophisticated reporters who are participants of Google's Trusted Copyright Removal Program,<sup>184</sup> rather than the “unsophisticated” users.<sup>185</sup> TCRP users are responsible for an increasing number of “no URI” complaints over the period between 2011 and 2015, going from issuing 26.4% of all “no URI” complaints in 2011 to 98.0% of all “no URI” complaints 2015, as the following graph below demonstrates.

---

<sup>183</sup> See, e.g., *Book DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (June 10, 2011) (on file with author) (stating that the reporter entered “DON’T KNOW WHAT THIS MEANS OR HWO [sic] TO ADDRESS IT”); see also, e.g., *eBooks DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Apr. 26, 2011) (on file with author) (stating that the reporter entered “Find it yourself, mother\*\*\*\*\*!” when asked to supply the “URL of the allegedly infringing material in our search results:”); see also, e.g., *DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Feb. 3, 2011) (on file with author) (stating “f\*\*\* that, your job to sort that” when similarly asked to supply the “URL of the allegedly infringing material in our search results:”).

<sup>184</sup> See Seng, *supra* note 19, at 414–18.

<sup>185</sup> Of course, trusted users can elect to serve Search notices on Google without signing in as trusted users. But this would be unlikely since there is no real advantage in doing so.



*Figure 5: Plot of Percentage of Notices and Complaints with No Takedown Requests (by TCRP Reporters) between 2011 and 2015<sup>186</sup>*

The following table breaks down these “empty request” complaints by the top reporters, including trusted users.

---

<sup>186</sup> See *supra* Table 4.

Reporter	Total Complaints Issued (A)	Total Notices Issued (B)	No. of Empty Request Complaints (C) (% of Total Complaints = (C)/(A))	No. of Notices with Empty Request Complaints (D) (% of Total Notices = (D)/(B))
1. BPI (British Recorded Music Industry) Ltd*	46,254,807	311,354	<b>2,791,927</b> <b>(6.04%)</b>	89,129 (28.6%)
2. Recording Industry Association of America, Inc.*	19,568,599	24,603	455,828 (2.33%)	14,625 (59.4%)
3. AudioLock.NET*	3,437,438	839,986	19,196 (0.56%)	12,458 (1.48%)
4. Mark Ayling*	2,211,693	5,485	4,073 (0.18%)	601 (11.0%)
5. Stichting BREIN*	1,943,589	19,741	3,950 (0.20%)	2,131 (10.8%)
6. Digimarc*	1,933,016	171,736	<b>109,674</b> <b>(5.67%)</b>	26,814 (15.6%)
7. Unidam, Inc.*	1,276,036	22,771	33,355 (2.61%)	4,967 (21.8%)
8. Link-Busters*	1,139,533	76,102	<b>197,600</b> <b>(17.3%)</b>	25,232 (33.2%)
9. MUSO.com Anti-piracy*	1,008,703	212,894	<b>80,180</b> <b>(7.95%)</b>	23,145 (10.9%)
10. Degban*	817,074	139,675	15,403 (1.89%)	4,453 (3.19%)
11. Vobile Inc*	631,370	34,838	<b>56,745</b> <b>(8.99%)</b>	8,866 (25.5%)
12. Counterfeit.Technology*	624,715	63,731	4,034 (0.65%)	1,898 (2.98%)
13. IP-Echelon Pty Ltd*	601,669	14,868	1,476 (0.25%)	513 (3.45%)
14. proMedia	570,285	77,502	2,638 (0.46%)	1,706 (2.20%)
15. Attributor*	423,910	18,838	19,716 (4.65%)	644 (3.42%)
16. Associação Antipirataria Cinema e Musica*	363,388	4,155	<b>134,452</b> <b>(37.0%)</b>	3,222 (77.6%)
17. Remove Your Media LLC*	354246	21247	970 (0.27%)	289 (1.36%)
18. CAPIE*	293,199	5,365	<b>122,982</b> <b>(41.9%)</b>	2,918 (54.4%)
19. The Publishers Association*	280960	11455	<b>62,034</b> <b>(22.1%)</b>	5,640 (49.2%)
20. APCM Mexico*	257,324	3,686	<b>158,178</b> <b>(61.5%)</b>	3,014 (81.8%)

Table 5: Top 20 Reporters by complaints, listing the total numbers of complaints, notices, empty request complaints and notices with empty request complaints between 2011 and 2015 ( $N_2=3,299,265$ ); TCRP members are marked with \*; those with 5% or more empty requests are boldfaced.<sup>187</sup>

The top transgressor who has sent out the most “empty request” complaints between 2011 and 2015 is BPI.<sup>188</sup> At over 2.7

<sup>187</sup> Results obtained by using MySQL to conduct on the second dataset a COUNT of notices and complaints that have either no requests or empty requests, sorted by the top 20 reporters, and combining that with a COUNT of the total number of notices and complaints for these reporters.

<sup>188</sup> See *supra* Table 5.

million complaints, this is over six times more than the next largest transgressor, RIAA, over five years.<sup>189</sup> And it has done so repeatedly, since almost 3 in 10 of its notices are notices containing complaints with empty requests.<sup>190</sup> In fact, as can be seen from the table, of the top twenty reporters who have issued the largest number of complaints between 2011 and 2015, almost half have 5% or more of empty request complaints, and *all* of these reporters fit the profile of TCRP “trusted” users.<sup>191</sup>

This looks less like an isolated incident and more like an industry phenomenon and calls for some explanation. One possible explanation<sup>192</sup> is that there were originally takedown requests associated with these “empty request” complaints, but they were “culled” by the reporter prior to submission.<sup>193</sup> The heavily-redacted

---

<sup>189</sup> See *supra* Table 5.

<sup>190</sup> See *supra* Table 5.

<sup>191</sup> The 5% cut-off is chosen because it is the most usual statistical significance level selected for rejecting the null hypothesis. As Lumen does not preserve information on whether a particular sender is a trusted user, suspected TCRP reporters are flagged as such based on their volume of complaints and requests that exceed the allowable limits in Google’s webforms that are available to the general public. (Currently the limits are up to 10 complaints per notice and up to 1,000 takedown requests per complaint: see *Copyright Removal*, GOOGLE, at [https://www.google.com/webmasters/tools/legal-removal-request?hl=en&pid=0&complaint\\_type=1](https://www.google.com/webmasters/tools/legal-removal-request?hl=en&pid=0&complaint_type=1) which permits up to 10 “groups” of complaints (see *supra* note 161), each of which relates to one copyright work, and permits the entry of up to 1000 infringing URLs (one URL per line)).

<sup>192</sup> Another possible explanation is that reporters were seeking to cast on the service providers “apparent knowledge” of possible infringing activity by identifying and thereby flagging the copyrighted works and leaving it to the service providers to locate the infringing materials. Such an explanation, while plausible, would be rejected by the courts. See *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102, 1113 (9th Cir. 2007) (holding as ineffective notices that separately supply the ownership information, the description of the work and the work location, on the basis that the DMCA notification procedures “place the burden of policing copyright infringement—identifying the potentially infringing material and adequately documenting infringement—squarely on the owners of the copyright” and that such notices will not cast any knowledge of infringement on the service providers because they are ineffective).

<sup>193</sup> See, e.g., *BPI DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Nov. 2, 2012) (on file with author) (making nine substantive complaints but putting in takedown requests for only two of these complaints); see also, e.g., *Music DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Dec. 18, 2012) (on file with author) (making ten complaints but only four come with takedown requests).

declaration of David Kaplan in *Disney Enterprises, Inc. v. Hotfile Corp.*<sup>194</sup> describes this as part of the decision making process on the part of reporters to avoid sending notices on misidentified content, whose effect is “a system that, by design, favors excluding [the take down of infringing] files rather than potentially misidentifying files.”<sup>195</sup>

While this is commendable, it also behooves reporters to completely remove these non-actionable complaints from their submitted notices. A failure to remove them may invite allegations that the good faith and accuracy attestations have been breached, and suggestions that TCRP users are cavalier in detecting infringement and respecting the sanctity of their own takedown complaints.

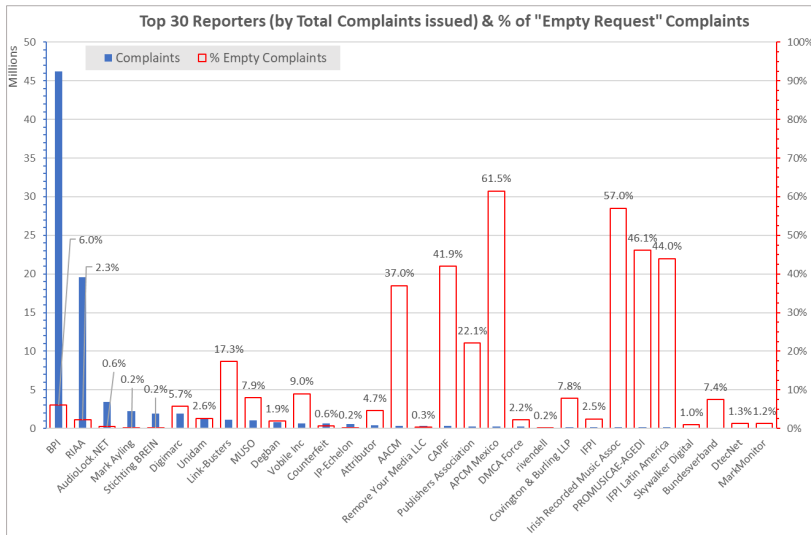


Figure 6: Top 30 Reporters by Total Complaints issued and % of “empty request” complaints between 2011 and 2015 ( $N_2=3,299,265$ )<sup>196</sup>

Lest it be contended that empty requests are unavoidable or that empty requests represent a trade-off between volume and precision, one must single out Stichting BREIN (0.20%),

<sup>194</sup> Declaration of David Kaplan in Support of Warner Bros. Entertainment’s Motion for Summary Judgment, at 5, *Disney Enter., Inc. v. Hotfile Corp.*, No 1:11-cv-20427-KMW (S.D. Fla. Feb. 28, 2012), ECF No. 308-1, <https://cases.justia.com/federal/district-courts/florida/flsdce/1:2011cv20427/373206/308/1.pdf?ts=1376979552>.

<sup>195</sup> *Id.*

<sup>196</sup> The chart is a graphical representation of Table 5, applying the same methodology but extending it to the top 30 reporters by total complaints.

AudioLock.NET (0.56%), Degban (1.89%) and RIAA (2.33%) for commendation.<sup>197</sup> Not only do they have the lowest ratio of empty request complaints to complaints among the reporters, they have done so while amongst the top 10 reporters of takedown complaints between 2011 to 2015 (at 1.94 million, 3.44 million, 0.82 million and 19.6 million complaints respectively), not far behind BPI (at 46.3 million complaints).<sup>198</sup> This is more than ample demonstration that empty request complaints are avoidable by reporters and can easily be fixed by introducing a simple adjustment to the reporters' systems prior to submission of the takedown requests.

### *E. Summary*

In sum, adopting a conservative test which narrowly equates a missing formality with statutory non-compliance, the above analysis shows that there is a sharp legal and empirical difference between the functional formalities and the non-functional formalities in notices. While non-functional formalities exhibit a low error rate of 0.02% (up to 2012),<sup>199</sup> functional formalities exhibit a much higher error rate that averages to 3.7% (for complaints with missing copyright work descriptions) and 5.5% (for complaints with empty takedown requests).<sup>200</sup> If a complaint exhibiting functional errors is considered to taint the parent notice, the error rates average to 5.5% (for notices with missing copyright work descriptions) and 9.8% (for notices with empty takedown requests).<sup>201</sup> It will be noted that the errors with the functional formalities, which are the critical elements of a takedown notice, are several orders of magnitude higher than the errors with non-functional formalities. Furthermore, while errors with non-functional formalities are on the decline,<sup>202</sup> owing to the use of form-based notices, errors with functional formalities have at best kept stable within the same range,<sup>203</sup> suggesting that the divergence between the

---

<sup>197</sup> See *supra* Figure 6.

<sup>198</sup> See *supra* Figure 6.

<sup>199</sup> See *supra* Table 3; see *supra* note 149.

<sup>200</sup> See *supra* Table 4; see *Websearch Infringement Notification via Online Form Complaint*, *supra* note 162.

<sup>201</sup> See *supra* Table 4; see *Websearch Infringement Notification via Online Form Complaint*, *supra* note 162.

<sup>202</sup> See *supra* the discussion immediately following Table 3; see *supra* note 150.

<sup>203</sup> See *supra* the discussion immediately following Table 4.

error rates of non-functional formalities and functional formalities has since grown further.

At one level, this difference can be explained as one of form (non-functional formalities) over substance (functional formalities), with the former going to the signatures and the prescribed attestations, and the latter going to the notice claims and takedown requests. As such, it is natural to expect the errors in functional formalities to be higher.<sup>204</sup>

That may be so. But even then, the *magnitude* of these functional errors still gives cause for concern. Copyright owners and reporters may argue that errors in takedown notices are an inevitable byproduct of enforcement, as they represent a tradeoff between accuracy and efficiency.<sup>205</sup> But inaccuracies manifesting themselves as errors also represent a cost: not just the service provider's cost of processing but also a cost to the copyright owner (and less so, for the reporter)<sup>206</sup> because they represent a missed enforcement opportunity. Also, it has to be remembered that in this part of the paper, we are only concerned with *formality* errors, and not even substantive errors. Since errors in the form of missing formalities are totally preventable at very low cost, errors such as those examined above should not even count as enforcement errors because they are really *process* errors—errors introduced in the administration and processing of takedown notices and requests. There is simply no-good reason why the same processes that have minimized (if not eliminated) the errors in non-functional formalities could not be applied to eliminate the same errors with functional formalities.

---

<sup>204</sup> See *supra* note 145 for a discussion about Google's form submission page which requires that the submitted notices be "checked" for compliance with the non-functional formalities such as attestations before the notice can be submitted. On the other hand, substantive information has to be supplied for the functional formalities. It should be noted that even though Google's Web Search takedown form (now) checks each submitted notice for the requisite reporter information, copyright work description, location of infringing material (URL) and signature fields (and declines to accept submissions without these fields), it should be questioned whether those notices with missing descriptions or missing requests were submitted through this interface (which would have disallowed such submissions) or through a different interface set up for TCRP reporters.

<sup>205</sup> See *e.g.*, *Disney Enterprises, Inc. v. Hotfile Corp.*, 2013 WL 6336286 at \*14 (S.D. Florida, 2013).

<sup>206</sup> There may be a question as to whether a reporter is paid by the owner if the takedown request is rejected by the service provider.



### III. SUBSTANTIVE ERRORS IN TAKEDOWN NOTICES

So far, the study has focused on an examination of errors associated with the failure to comply with the functional and non-functional formalities in the DMCA. If formality errors taint at least 5.5% of takedown notices, what proportion of substantive errors—errors that raise substantive legal questions which go to the underlying claim for alleged copyright infringement—will afflict the notices? In the Urban and Quilter study, the authors found that at least one third of all notices (N=876) manually reviewed were erroneous because the claims were over non-copyrightable subject matter or raised fair use and other substantive defenses,<sup>207</sup> a statistic that remained relatively stable in the updated study.<sup>208</sup> In the decade after the passage of the DMCA, has the situation improved?

On closer examination of the issue, it turns out that in order to assess if there are substantive errors in a takedown notice, a detailed evaluation of the notice contents as well as the targeted hosting page is required. However, since 2008, the use of online forms and “robotakedowns” has prevailed.<sup>209</sup> With substantial investments into systems for processing these takedowns by service providers,<sup>210</sup> this has paradoxically shortened turnaround times for responding to takedown notices and their requests.<sup>211</sup> For instance, Google indicates that it responds to takedown requests within 24 hours of their submission.<sup>212</sup> And generally, cyberlocker and other web hosting sites respond to such requests in a matter of days.<sup>213</sup> When the takedown notice is eventually

---

<sup>207</sup> Urban & Quilter, *supra* note 30, at 666–67.

<sup>208</sup> Urban et al., *supra* note 33, at 88.

<sup>209</sup> See Seng, *supra* note 19, at 398–400.

<sup>210</sup> Section 512 Study, *supra* note 17, at 81865.

<sup>211</sup> See e.g., Kent Walker, *Making Copyright Work Better Online*, GOOGLE (Dec. 2, 2010), <https://publicpolicy.googleblog.com/2010/12/making-copyright-work-better-online.html> (noting that Google would build tools to improve and shorten the time for processing of takedown notices).

<sup>212</sup> See, e.g., *FAQ – Google Transparency Report*, GOOGLE, [http://www.google.com/transparencyreport/removals/copyright/faq/#response\\_time](http://www.google.com/transparencyreport/removals/copyright/faq/#response_time) (last visited May 11, 2015) (noting that as of December 2012, the average processing time for all removal requests submitted via web forms for Search is approximately 6 hours).

<sup>213</sup> *The DMCA is Broken...*, THE TRICHORDIST (July 18, 2012), <http://thetrichordist.com/2012/07/18/the-dmca-is-broken/> (noting that most

published on Chilling Effects or Lumen about 1 week later, it is more often the case that both the targeted link (takedown request) and the hosting page have been disabled. These two reasons together make it infeasible to examine and evaluate each notice for its underlying substantive copyright claims.<sup>214</sup>

To this end, in this study, it is proposed to evaluate the notices for their underlying substantive copyright claims *indirectly* by focusing on two narrow issues. The first is that an effective notice has to properly identify the copyright work *and* the copyright owner or exclusive licensee of rights in that work. After all, the statement of the reporter's

---

cyberlocker sites remove the allegedly offending material between 24 and 48 hours); *How to File DMCA Takedown Notice*, PIRACYTAKEDOWN (May 1, 2014), <http://piracytakedown.com/blog/dmca-takedown-notice> (noting that most cyberlockers will remove infringing content in less than 2 days after receiving takedown notices); Tobias Lauinger et al., *Clickonomics: Determining the Effect of Anti-Piracy Measures for One-Click Hosting*, PROCEEDINGS OF NDSS SYMPOSIUM 2013 (2013) (empirically finding that most cyberlockers lapse old content after 30 days) [hereinafter *Clickonomics*]. For a more extreme example, the cyberlocker site Hotfile gave a content provider, Warner Bros., access to a “takedown tool” which allowed Warner as a trusted content owner to access Hotfile’s system to identify and automatically remove offending links without any action by Hotfile. *See* Disney Enter., Inc. v. Hotfile Corp., No. 11-20427-CIV, 2013 WL 6336286, at 11 (S.D. Fla. Sept. 20, 2013).

<sup>214</sup> In the Urban updated study, the investigation into substantive errors was conducted by an ex-post manual quantitative examination of a random sample of 1,827 Lumen takedown notices issued between May and October 2013, by reviewing the notices against Google Web Search index entries and Google Image Services thumbnails. The issue with this methodology is that given the lapse of time (the review itself took place most likely at least after February 2014 (Urban et al., *supra* note 33, at 78 n.209), this examination can only be conducted on online resources that have remained accessible and available *at least 4 months after* the takedowns were issued. These will tend to be resources whose uses are *not* illicit or are authorised fair use. Therefore, the availability of the resource itself for review into substantive errors creates a selection bias known as *survivorship bias* that skews the ensuing investigation and analysis in favour of the finding that there is a *higher* incidence of substantive errors (since no investigations can be conducted into resources that are presumptively illicit or unlicensed and have been removed). The findings of the Urban updated study therefore represent the *higher* bound of the extent of substantive errors in takedown notices. For an explanation of survivorship bias, *see* Katy Milkman, *The Perils of “Survivorship Bias”*, SCIENTIFIC AMERICAN (Feb. 11, 2020), <https://www.scientificamerican.com/article/the-perils-of-survivorship-bias/> (last visited Feb 8, 2021).

good faith belief is that the “use of the [described copyright] *material* in the manner complained of is not authorized by the *copyright owner*, its *agent*, or the law.”<sup>215</sup> The reporter also attests to the fact that “the information in the notification is *accurate*.”<sup>216</sup> The second is that an effective takedown request has to refer to an ostensibly valid allegation of infringing activity, consistently with the reporter’s attested statement that the “use of the material in the manner complained of *not authorized*,”<sup>217</sup> and that “the information in the notification is *accurate*”<sup>218</sup> and that the reporter’s notification is targeting an exclusive right that is *allegedly infringed*.”<sup>219</sup> These two issues are certainly not exhaustive of all the substantive legal issues relating to a takedown notice. But, as will be shown later, they are tractable, given the large volumes of notices in the dataset, and their results are illuminative.

#### A. Notices that Misidentify the Copyright Owner

There has been anecdotal evidence of the numerous instances where reporters have wrongly claimed on behalf of copyright owners’ infringements over works to which they do not own copyright.<sup>220</sup> Such evidence has ceased to be purely anecdotal. In the *Disney Enterprises, Inc. v. Hotfile Corp.* litigation, it was successfully alleged that Warner Brothers as a content provider had (apparently) sought to remove files that undisputedly belonged to Electronic Arts, Inc.,<sup>221</sup> and had (intentionally) sought to remove a free software program, JDownloader, that was created by a German company, for which Warner does not own or have rights to.<sup>222</sup> In the *Hotfile* litigation, the scale of such infractions was small—the former involved 271 requests and the latter involved 8 requests.<sup>223</sup> Nonetheless, they sufficed to enable Hotfile to maintain its counter-claim against Warner Brothers for misrepresentation under the DMCA; with the court ruling that

<sup>215</sup> 17 U.S.C. § 512(c)(3)(A)(v) (emphasis added).

<sup>216</sup> 17 U.S.C. § 512(c)(3)(A)(vi) (emphasis added).

<sup>217</sup> 17 U.S.C. § 512(c)(3)(A)(v) (emphasis added).

<sup>218</sup> 17 U.S.C. § 512(c)(3)(A)(vi) (emphasis added).

<sup>219</sup> *Id.*

<sup>220</sup> See *supra* notes 38–41 and accompanying text.

<sup>221</sup> *Disney Enter., Inc. v. Hotfile Corp.*, No. 11-20427-CIV, 2013 WL 6336286, at 15 (S.D. Fla. Sept. 20, 2013) (noting that Electronic Arts, Inc. subsequently gave Warner after-the-fact permission to remove the files).

<sup>222</sup> *Id.*

<sup>223</sup> *Id.* at 17.

“there is sufficient evidence in the record to suggest that Warner intentionally targeted files it knew it had no right to remove.”<sup>224</sup>

In the course of conducting this study on the first dataset through the Chilling Effects database,<sup>225</sup> at least two more instances of infractions of the same type were found, albeit on a much larger scale.

In the first instance, NBCUniversal’s takedown notices wrongly described one “B\*\*\*\*\* P\*\*\*\*\*” as the copyright owner in its notices. As a result, a total of 169 notices, comprising a total of 132,299 requests, were served on Google between August and December 2011.<sup>226</sup> A summary of these erroneous notices and requests follows:

Month/Year	No. of Notices	Total No. of Requests Submitted	Total No. of Requests Rejected	DMCA Section	Identified Copyright Owner	Actual Copyright Owner (presumed)	Reporter	Google Service
Aug. 2 to Aug. 30, 2011	49	37,072	889	s512(d)	B***** P*****	NBCUniversal	NBCUniversal	Search
Sep. 1 to Sep. 28, 2011	42	36,803	998	s512(d)	B***** P*****	NBCUniversal	NBCUniversal	Search
Oct. 2 to Oct. 11, 2011	30	19,818	272	s512(d)	B***** P*****	NBCUniversal	NBCUniversal	Search
Nov. 1 to Nov. 22, 2011	40	32,636	968	s512(d)	B***** P*****	NBCUniversal	NBCUniversal	Search
Dec. 6 to Dec. 8, 2011	8	5,970	4	s512(d)	B***** P*****	NBCUniversal	NBCUniversal	Search

Table 6: NBCUniversal's Erroneous Takedown Requests<sup>227</sup>

As the reporter, NBCUniversal had attested in the notices that the 132,299 requests covered in the table above pertained to the use of infringing materials which were not authorized by the copyright owner “B\*\*\*\*\* P\*\*\*\*\*” and that it had authority from “B\*\*\*\*\* P\*\*\*\*\*” to act to remove these requests.<sup>228</sup> It would not be apparent

<sup>224</sup> *Id.* at 48.

<sup>225</sup> See generally LUMEN, *supra* note 54. During a verification of the above results with the Lumen database, an interesting divergence was noted. While the number of notices misidentifying “B\*\*\*\*\* P\*\*\*\*\*” as a copyright owner had remained similar (with 250 notices containing 188,488 complaints), the number of notices misidentifying “P\*\*\*\*\* T\*\*\*\*\*” had dropped drastically to 1 notice. The reasons for the discrepancy between the Chilling Effects and Lumen databases are currently unknown.

<sup>226</sup> See *infra* Table 6.

<sup>227</sup> Results obtained by using MySQL to conduct on the first dataset a search for “B\*\*\*\*\* P\*\*\*\*\*” in the contents of copy of the original, complete and unredacted notice.

<sup>228</sup> These are standard attestations made in every takedown request. See 17 U.S.C. §§ 512(c)(3)(A)(v)–(vi).

to a reviewer of these notices that “B\*\*\*\*\* P\*\*\*\*\*” was one of NBCUniversal’s employees.<sup>229</sup> Of these, Google only rejected 3,131 of these takedown requests (a rejection rate of only 2.37%).<sup>230</sup>

In the second instance, it was found, on the first dataset, that the reporting agent Marketly.com, which has been described as an Internet agent set up by an ex-Microsoft employee “P\*\*\*\* T\*\*\*\*\*” to manage Microsoft’s anti-piracy program,<sup>231</sup> described himself in no less than 1,114 notices as the copyright owner of Microsoft software and games in its takedown notices between June and September 2011.<sup>232</sup> Again, as above, there was no reason to associate the ex-Microsoft employee with ownership of the copyright in Microsoft software and games. The numbers of affected requests are as follows:

Month/ Year	No. of Notices	Total No. of Requests Submitted	Total No. of Requests Rejected	DMCA Section	Identified Copyright Owner	Actual Copy-right Owner (presumed)	Reporter	Google Service
Jun. 13 to Jun. 30, 2011	23	13,447	951	s512(d)	P**** T*****	Microsoft Corp.	<a href="#">Marketly.com</a>	Search
Jul. 1 to Jul. 31, 2011	244	112,970	504	s512(d)	P**** T*****	Microsoft Corp.	<a href="#">Marketly.com</a>	Search
Aug. 1 to Aug. 31, 2011	393	32,714	827	s512(d)	P**** T*****	Microsoft Corp.	<a href="#">Marketly.com</a>	Search
Sep. 1 to Sep. 26, 2011	457	100,600	668	s512(d)	P**** T*****	Microsoft Corp.	<a href="#">Marketly.com</a>	Search

Table 7: Marketly.com's Erroneous Takedown Requests<sup>233</sup>

All in, more than ¼ million takedown requests were issued under these erroneous notices.<sup>234</sup> Of these 259,731 requests, Google

<sup>229</sup> A LinkedIn search verified the relationship between “B\*\*\*\*\* P\*\*\*\*\*” and NBCUniversal. The link to the LinkedIn search for “B\*\*\*\*\* P\*\*\*\*\*” would not be supplied in order to protect the personal identity of this individual.

<sup>230</sup> See *supra* Table 6.

<sup>231</sup> Cyrus Farivar, *Microsoft Outsources Copyright Enforcement to Small Redmond Company*, ARSTECHNICA (May 29, 2012), <http://arstechnica.com/business/2012/05/microsoft-outsources-copyright-enforcement-to-small-redmond-company/> (stating that Marketly was founded by a former Microsoft employee with no legal background).

<sup>232</sup> Results obtained by using MySQL to conduct on the first dataset a search for “P\*\*\*\* T\*\*\*\*\*” in the contents of copy of the original, complete and unredacted notice.

<sup>233</sup> *Id.*

<sup>234</sup> See *supra* Table 7.

only rejected 2,950 of them, which works out to a rejection rate of only 1.14%.<sup>235</sup>

For reasons explained above,<sup>236</sup> these notices amounted to potential misrepresentations on the part of the reporter and are actionable by the recipient service provider. These are not correctable, non-functional formalities errors which are described in the earlier part of this paper.<sup>237</sup> A service provider can mount a possibly valid challenge that these notices are inaccurate, that the mistakes made in these notices amount to a substantive error on the part of NBCUniversal, Marketly.com, and BPI, and as a consequence, the service provider ought to have rejected *all* of the notices.

This study did not specifically set out to locate these types of errors. And from one perspective, it could be argued that the *scale* of these errors is small—only 380,379 requests out of 56,991,045 (or 0.67%) were detected in the first dataset.<sup>238</sup> But what is alarming is the *magnitude, frequency and systematic nature* of these errors, which remained undetected and the erroneous notices repeatedly recycled for months on end. While we may excuse these errors on the basis that they arose from programs that are misconfigured with wrong information, automated systems propagated these errors across hundreds and thousands of takedown requests.<sup>239</sup> All these seem to evidence a “configure and forget” approach on the part of the reporters, an absence of manual review, and a lack of rigorous oversight of the entire takedown process and its aftermath.

A substantive error committed on arguably an even larger scale can be seen in what this study calls the “Megaupload test.”

### B. *The Megaupload Test*

When Megaupload and its subsidiary sites were shut down by the U.S. government on January 19, 2012, following the indictment and

---

<sup>235</sup> See *supra* Table 7.

<sup>236</sup> See *supra* the discussion accompanying notes 169–177.

<sup>237</sup> See *supra* the discussion accompanying notes 109–118.

<sup>238</sup> This is the size of the original Chilling Effects dataset examined between 2009 and 2012, the first dataset. See Seng, *supra* note 19, at 383.

<sup>239</sup> Another example of such an automated propagation error is where there is a hundred-fold repetition of the *same* description of the *same* copyright complaint in one takedown notice. See, e.g., *BPI DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (May 4, 2012) (on file with author).

arrests of its owners for violating copyright laws,<sup>240</sup> it came as a real shock to the hosting industry or cyberlockers.<sup>241</sup> Other file hosting sites took almost immediate steps to either limit the functionality of their services, such as withdrawing the ability to allow their subscribers to share links to uploaded files, or they shut down completely.<sup>242</sup> By that single operation, not only had the music and movie industries pulled the plug on what was alleged to be a huge source of unlicensed materials, they had also irrevocably changed the face of the hosting industry.<sup>243</sup> While some say that this has had little or no effect on piracy because other hosting companies were quick to take the place of those that shut down,<sup>244</sup> other commentators noted that this has forced the cyberlocker industry to clean up its act, including scaling back or cancelling its affiliate programs.<sup>245</sup> Search engines like Google

---

<sup>240</sup> Jacob Ganz & Laura Sydell, *Megaupload Shut Down by the FBI*, NPR (Jan. 19, 2012), <https://www.npr.org/sections/therecord/2012/01/20/145474712/megaupload-shut-down-by-the-fbi>.

<sup>241</sup> See, e.g., Cecilia Kang, *Megaupload Shutdown Raises New Internet-sharing Fears*, THE WASHINGTON POST (Jan. 20, 2012), [https://www.washingtonpost.com/business/technology/megaupload-shutdown-raises-new-internet-sharing-fears/2012/01/20/gIQATHRtEQ\\_story.html](https://www.washingtonpost.com/business/technology/megaupload-shutdown-raises-new-internet-sharing-fears/2012/01/20/gIQATHRtEQ_story.html).

<sup>242</sup> See, e.g., Sami Yenigun, *Other File-Sharing Sites: "We're Not Megaupload"*, NPR (Jan 27, 2012), <https://www.npr.org/2012/01/27/145919516/other-file-sharing-sites-were-not-megaupload>; John Plunkett, *BTjunkie "Voluntarily" Shuts Down*, THE GUARDIAN (Feb. 6, 2012), <https://www.theguardian.com/technology/2012/feb/06/btjunkie-voluntarily-shuts-down#:~:text=The%20operators%20of%20BTjunkie%2C%20which,it%27s%20time%20to%20move%20on.>

<sup>243</sup> Ernesto Van der Sar, *MPAA: Megaupload Shutdown was Massive Success*, TORRENTFREAK (Dec. 5, 2012), <http://torrentfreak.com/mpaa-megaupload-shutdown-was-massive-success-121205/>.

<sup>244</sup> See, e.g., Elias Groll, *The Feds Brought Down the World's Biggest File-Sharing Site*, FOREIGN POLICY (Jul. 22, 2016), <https://foreignpolicy.com/2016/07/22/the-feds-brought-down-the-worlds-biggest-file-sharing-site/>.

<sup>245</sup> See, e.g., Sami Yenigun, *Other File-Sharing Sites: "We're Not Megaupload"*, NPR (Jan. 27, 2012), <https://www.npr.org/2012/01/27/145919516/other-file-sharing-sites-were-not-megaupload>.

voluntarily cancelled their search results for megaupload.com.<sup>246</sup> And, it has been claimed that, as a result, sales of licensed digital content improved.<sup>247</sup>

Yet, more than a year after the seizure of Megaupload, reporting agents are still submitting takedown requests for Megaupload links.<sup>248</sup> And, it is not just Megaupload: the data shows that other one-click hosting companies, which have since shut down, are still targeted through the takedown notices.

The following table shows the status of all the Megaupload sites and various other shuttered one-click sites and hosts after the Megaupload incident. As of the time of writing, as well as further tests in 2019 and 2021, the sites remain inaccessible:

Hosting Site	Date of Action	Action
<b>megaupload.com</b>	Jan. 19, 2012	Closed
<b>megapix.com</b>	Jan. 19, 2012	Closed
<b>megavideo.com</b>	Jan. 19, 2012	Closed
<b>megalive.com</b>	Jan. 19, 2012	Closed
<b>megabox.com</b>	Jan. 19, 2012	Closed
<b>megaporn.com</b>	Jan. 19, 2012	Closed
<b>megarotic.com</b>	Jan. 19, 2012	Closed
<b>sexuploader.com</b>	Jan. 19, 2012	Closed

---

<sup>246</sup> Ernesto Van der Sar, *Google Removes 'BitTorrent' from Piracy Search Filter*, TORRENTFREAK (Sept. 24, 2013), <http://torrentfreak.com/google-removes-bittorrent-from-piracy-search-filter-130924/>.

<sup>247</sup> See Brett Danaher & Michael D. Smith, *Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales*, SSRN (Sept. 14, 2013), <http://ssrn.com/abstract=2229349>.

<sup>248</sup> Mike Masnick, *Copyright Holders Still Sending DMCA Takedowns on Content That's Been Gone for Months*, TECHDIRT (Sept. 7, 2012), <http://www.techdirt.com/articles/20120907/13121720312/copyright-holders-still-sending-dmca-takedowns-content-thats-been-gone-months.shtml>; *Anti-Piracy Outfits Think Megaupload, Demonoid & BTjunkie Are Still Alive*, TORRENTFREAK (Sept. 7, 2012), <http://torrentfreak.com/anti-piracy-outfits-think-megaupload-demonoid-btjunkie-are-still-alive-120907/>.



Hosting Site	Date of Action	Action
<b>filesonic.com</b>	Jan. 22, 2012 Sep. 3, 2012	Accounts and files deleted; file sharing capabilities disabled; <sup>249</sup> closed <sup>250</sup>
<b>fileserve.com</b>	Jan. 22, 2012	Affiliates program withdrawn; accounts deleted; file sharing capabilities disabled; <sup>251</sup> DMCA complaint page not working <sup>252</sup>
<b>filejungle.com</b>	Jan. 23, 2012	Affiliates program withdrawn; file sharing capabilities disabled <sup>253</sup>
<b>uploadstation.com</b>	Jan. 23, 2012	File sharing capabilities disabled <sup>254</sup>

<sup>249</sup> Casey Chan, *FileSonic Just Killed Itself by Disabling File Sharing*, GIZMODO (Jan. 22, 2012), <http://gizmodo.com/5878287/filesonic-just-killed-itself-by-disabling-file-sharing>; Steven Musil, *FileSonic Disables File Sharing in Wake of MegaUpload Arrests*, CNET (Jan. 22, 2012), [http://news.cnet.com/8301-1023\\_3-57363594-93/filesonic-disables-file-sharing-in-wake-of-megaupload-arrests/](http://news.cnet.com/8301-1023_3-57363594-93/filesonic-disables-file-sharing-in-wake-of-megaupload-arrests/).

<sup>250</sup> Andre Yoskowitz, *FileSonic Finally Shut Down for Good?*, AFTERDAWN (Sept. 3, 2012, 3:38 PM), [http://www.afterdawn.com/news/article.cfm/2012/09/03/filesonic\\_finally\\_shut\\_down\\_for\\_good](http://www.afterdawn.com/news/article.cfm/2012/09/03/filesonic_finally_shut_down_for_good); Gunner Thorne, *Filesonic.com is Offline*, ONLINE FILE STORAGE (Feb. 4, 2016), <https://web.archive.org/web/20160821095300/http://onlinefilestorage.com/filesonic-com-is-offline-2129>.

<sup>251</sup> Andy Maxwell, *Cyberlocker Ecosystem Shocked as Big Players Take Drastic Action*, TORRENTFREAK (Jan. 23, 2012), <http://torrentfreak.com/cyberlocker-ecosystem-shocked-as-big-players-take-drastic-action-120123/>.

<sup>252</sup> DMCA, FILESERVE, <http://dmca.fileserve.com/login.php> (last visited Mar. 12, 2013).

<sup>253</sup> Maxwell, *supra* note 251.

<sup>254</sup> Raptile, *UploadStation.com Removes it's [sic] Affiliate Program*, WJUNCTION.COM (Jan. 23, 2012), <https://www.wjunction.com/threads/uploadstation-com-removes-its-affiliate-program.125422/>; UPLOADSTATION.COM, <http://uploadstation.com/> (since defunct, although comparisons of uploadstation.com's website on Wayback

Hosting Site	Date of Action	Action
<b>x7.to</b>	Jan. 23, 2012	Closed <sup>255</sup>
<b>uploadbox.com</b>	Jan. 30, 2012	Closed <sup>256</sup>
<b>btjunkie.org</b>	Feb. 5, 2012	Closed <sup>257</sup>
<b>quicksilverscreen.com</b>	Feb. 7, 2012	Closed
<b>wupload.com</b>	Apr. 3, 2012	File sharing capabilities disabled; <sup>258</sup> domain name is inactive—apparently closed
<b>demonoid.me</b>	Jul. 25, 2012	Taken down; <sup>259</sup> since closed <sup>260</sup>

*Table 8: List of Sites Closed After the Megaupload Takedown*

Mapping this information against all the takedown requests received between January 2011 and December 2015 in the second dataset yields the following results:

---

Machine immediately before and after WJunction’s article shows the removal of monetization links such as “Make Money,” “Redeem,” “Linking Tools,” and “Advertising.”); *see also* Maxwell, *supra* note 251.

<sup>255</sup> *x7.to Closed*, WJUNCTION.COM (Jan. 27, 2012),

<https://www.wjunction.com/threads/x7-to-closed.126704/>.

<sup>256</sup> Maxwell, *supra* note 251.

<sup>257</sup> *See* Plunkett, *supra* note 242.

<sup>258</sup> Andy Maxwell, *FileServe and Wupload Exit the File-Sharing Business*, TORRENTFREAK (Apr. 3, 2012), <http://torrentfreak.com/files-serve-and-wupload-exit-the-file-sharing-business-120403/>; Andy Maxwell, *We’re No Rogue Site: PutLocker Responds to Hollywood*, TORRENTFREAK (Apr. 3, 2012), <http://torrentfreak.com/were-no-rogue-site-putlocker-responds-to-hollywood-120403/>.

<sup>259</sup> Emil Protalinski, *Demonoid Busted by the Police*, ZDNET (Aug. 6, 2012, 9:38 PM), <http://www.zdnet.com/demonoid-busted-by-the-police-7000002208/>.

<sup>260</sup> Ernesto Van der Sar, *Demonoid Tracker Goes Down, Again*, TORRENTFREAK (Dec. 15, 2012), <http://torrentfreak.com/demonoid-tracker-goes-down-again-121215/>.

Site/ Domain Name	Closure/ Action Date (T)	All requests from T+0 to T+1 month	All requests from T+1 to T+2 months	All requests from T+2 to T+3 months	All requests from T+3 to T+6 months	All requests from T+6 to T+12 months	All requests from T+12 to T+24 months	All requests from T+12 to T+24 months
<b>megaupload</b>	Jan. 19, 2012	59	135	50	565	4,910	1,397	348
<b>megapix</b>	Jan. 19, 2012							
<b>megavideo</b>	Jan. 19, 2012					56	110	7
<b>megalive</b>	Jan. 19, 2012							
<b>megabox</b>	Jan. 19, 2012							
<b>megaporn</b>	Jan. 19, 2012					2	1	
<b>megarotic</b>	Jan. 19, 2012				1		1	
<b>sexuploader</b>	Jan. 19, 2012							3
<b>filejungle</b>	Jan. 23, 2012	951	5	2	3,005	70,981	45,088	105,559
<b>uploadstation</b>	Jan. 23, 2012	412	4,606	2,243	2,538	71,677	96,579	78,315
<b>x7.to</b>	Jan. 23, 2012			1	781	3,155	2,371	89
<b>uploadbox</b>	Jan. 30, 2012	3	3		294	190	13	725
<b>btjunkie</b>	Feb. 6, 2012	2,654	902	547	607	21,459	20,428	8,591
<b>quicksilver- screen</b>	Feb. 7, 2012					12,879	30	14
<b>fileserve</b>	Apr. 1, 2012	21	5,167	8,020	19,925	221,139	186,107	462,218
<b>wupload</b>	Apr. 3, 2012	8,474	10,830	11,094	18,617	162,692	86,263	218,483
<b>demonoid</b>	Jul. 25, 2012	300	319	205	481	2,367	241	440
<b>filesonic</b>	Sep. 3, 2012	5,523	2,781	5,975	133,745	40,486	468,207	165,848

Table 9: Closed or File-sharing Disabled Sites and their “Spent” Takedown Requests<sup>261</sup>

All in, using a generous 90-day grace period from the date of closure of the site, 2.74 million takedown requests were served on these disabled sites up to December 2015, not an insubstantial number.<sup>262</sup> The information here is better viewed in the following time plot that contrasts the takedown requests before and after the key closure/action date of January 19, 2012.

<sup>261</sup> Results obtained by way of a MySQL COUNT of second dataset of all requests whose domain names are as indicated, separated into the indicated time bands based on the date received of the notice where each request was found. As of 2015, shaded sites/domain names were closed; unshaded sites/domain names were still active but had disabled file sharing.

<sup>262</sup> See *supra* Table 9.



Figure 7: Top-10 “Spent” Takedown Requests by Domain Name between July 2011 and Dec. 2015 (weekly figures) (with shut down dates indicated against their respective domain names)<sup>263</sup>

<sup>263</sup> This chart is a graphical representation of Table 9; see *supra* Table 9.

From these tables and plot,<sup>264</sup> several rather surprising observations may be made. First, notwithstanding all the concerns by industry about how much piracy is being propagated by some of these high profile cyberlockers such as megalive, megabox, megaporn, megarotic and uploadbox, few or no takedown requests appear to be targeting them, both before as well as after January 19, 2012, or at least through Google's takedown system in the form of notices reported in Lumen.<sup>265</sup> Even if it were contended that the takedown requests were addressed directly to these hosts, it appears incongruous for the content providers and reporters to not make a request to remove any links that the Google search engine may have to the same content hosted by these sites.<sup>266</sup>

However, this does not square with the second observation, that although all these sites (and their file sharing features) are disabled, the data shows that large numbers of takedown requests targeting these sites are still directed through Google. In some instances, with sites like filesonic, fileserve and wupload, takedown requests *spiked* after the site has been made completely inaccessible.<sup>267</sup> Perhaps Google still retains cached copies of these infringing links. However, if these "dead" links do not enable the user to access any infringing material, surely retaining these links themselves does not constitute infringing activity.<sup>268</sup> What is even more extraordinary is that in some cases, such as that for Megaupload, Google has taken down *all* its own links in Google Search to Megaupload, when the site itself was seized by the U.S. government

---

<sup>264</sup> See *supra* Table 9; see *supra* Figure 7.

<sup>265</sup> See *supra* Table 9; see *supra* Figure 7.

<sup>266</sup> Most reporters will advise content providers to serve takedown notices on both cyberlockers and search engines. See *How to File DMCA Takedown Notice*, PIRACYTAKEDOWN (May 1, 2014), <http://piracytakedown.com/blog/dmca-takedown-notice>.

<sup>267</sup> See *supra* Table 9.

<sup>268</sup> If the links themselves do not constitute any infringing activity, identifying and targeting them in takedown requests is a breach of 17 U.S.C. § 512(c)(3)(A)(iii), which requires the reporters to *identify* the material "that is claimed to be infringing or to be the subject of infringing activity," to which the reporters have attested they have a "good faith belief" that such an activity is "not authorized by the copyright owner." 17 U.S.C. § 512(c)(3)(A)(v). This calls into question both whether they have actually identified the infringing activity and their good faith belief of their identification.

in January 2012.<sup>269</sup> Yet, owners and reporters continue to serve thousands of requests concerning Megaupload.<sup>270</sup>

---

<sup>269</sup> See Cask J. Thompson, *[SOPA OPERA] The Online War—The Internet Reacts Not-So-Nicely to Megaupload Shutdown.*, WORDSWITHMEANING.ORG (Jan. 20, 2012), archived at [http://web.archive.org/web/20120125211515if\\_/http://wordswithmeaning.org/2012/01/the-online-warthe-internet-reacts-not-so-nicely-to-megaupload-shutdown#axzz1kVUiEAS7](http://web.archive.org/web/20120125211515if_/http://wordswithmeaning.org/2012/01/the-online-warthe-internet-reacts-not-so-nicely-to-megaupload-shutdown#axzz1kVUiEAS7).

<sup>270</sup> As of May 2015, the Lumen database shows that takedown requests are still being directed at the defunct megavideo.com and x7.to domains. *See, e.g., DMCA (Copyright) Complaint to Google*, LUMEN (May 22, 2015), <https://www.lumendatabase.org/notices/10794321>.

Rank	Reporter	Total Requests	Total Requests that Fail the Megaupload test
1.	Degban	201,096,948	38,226
2.	BPI (British Recorded Music Industry) Ltd*	167,305,587	10,898
3.	rivendell	92,020,862	0
4.	Recording Industry Association of America, Inc.*	69,324,907	291
5.	DtecNet	57,534,056	1,078,219
6.	MarkMonitor AntiPiracy	55,538,930	389,604
7.	Remove Your Media LLC	55,097,505	539
8.	Takedown Piracy LLC	49,151,664	21,193
9.	AudioLock.NET	38,961,320	0
10.	Unidam, Inc.	29,730,630	300,225
11.	Skywalker Digital, Ltd.	27,417,459	33,988
12.	DMCA Force	24,009,736	1,017
13.	MUSO.com Anti-piracy	21,836,365	3,115
14.	IP-Echelon Pty Ltd	19,539,217	34
15.	Fox Group Legal*	17,907,645	0
16.	NBCUniversal*	17,370,446	333,668
17.	Walt Disney Company*	13,997,165	220
18.	Entura International Ltd	13,776,571	0
19.	xTakedowns	13,103,134	0
20.	Stichting BREIN*	12,596,190	0
21.	Link-Busters.com	11,208,365	48,755
22.	Marketly.com	11,150,221	906
23.	Removeyourcontent LLC	9,533,886	0
24.	APDIF do Brasil*	9,026,362	0
25.	Mark_Ayling	8,781,121	0
26.	APDIF – Mexico*	6,959,517	0
27.	Topple Track	6,387,034	241
28.	Digimarc	6,303,232	1,063
29.	Vobile Inc	5,725,773	1,165
30.	APCM México*	4,925,087	0

Table 10: Requests from Top-30 Reporters, showing Total Requests that Fail the Megaupload test; copyright owners and industry groups marked with \*<sup>271</sup>

Even if one assumes that reporting agents need some time to “catch-up” with their infringement detection and process these takedown requests, it would not explain why up to 90 days after these sites are closed, takedown requests are still being issued against these

---

<sup>271</sup> Results obtained by way of a MySQL COUNT of second dataset of all requests, and requests that target the disabled sites, identifying the reporter who sent them and grouping and sorting the reporters in descending order of the number of all requests; reporters identified as copyright owners and industry groups by using the Lumen database. *See Lumen – About Us, supra* note 53.

sites.<sup>272</sup> In fact, comparing the *before* and *after* closure takedown requests, it is patently clear that in every case, *more* requests were issued against these sites *after* they were closed.<sup>273</sup> Furthermore, these erroneous requests were not limited to a few reporting agents, but were distributed rather evenly among reporting agents, copyright owners and their industry groups.<sup>274</sup> That this is a totally avoidable error can be seen by noting that the takedown requests of some of the top reporters—APCM México, APDIF – Mexico, APDIF do Brasil, AudioLock.NET, Entura International Ltd, Fox Group Legal, Mark\_Ayling, Removeyourcontent LLC, rivendell, Stichting BREIN and xTakedowns, *pass* the Megaupload test.<sup>275</sup> In other words, unlike the other reporters, they sent out *not* a single takedown request to any of the closed sites.

All in, 2.74 million clearly invalid takedown requests to which every issuing reporter has attested to their accuracy were issued.<sup>276</sup> Although these amounted to only 0.23% of all takedown requests issued between 2011 and 2015,<sup>277</sup> based on the sheer frequency and repetitiveness of these errors, each and every one of these non-actionable requests is a knowing misrepresentation<sup>278</sup> that the reporters “should have known [about and not issued] if [they have] acted with reasonable care or diligence or would have had no substantial doubt had it been acting in good faith.”<sup>279</sup>

The most plausible (and most enlightening) explanation for this serious problem is that the owners and reporting agents are looking for infringing materials online by merely checking for the presence of certain search terms on third party sites that collect links to such content—aggregator sites—without verifying that the links actually work. As one owner/reporter asserted, given the bandwidth and processing limitations, it was not practical for the owners and reporters to download every allegedly infringing file before issuing a takedown

---

<sup>272</sup> See *supra* Figure 7.

<sup>273</sup> See *supra* Figure 7.

<sup>274</sup> See *supra* Table 10.

<sup>275</sup> See *supra* Table 10.

<sup>276</sup> See *supra* Table 9.

<sup>277</sup> A total of 1,169,359,565 DMCA takedown requests directed at Google are recorded between 2011 and 2015. See *supra* Table 1.

<sup>278</sup> 17 U.S.C. § 512(f).

<sup>279</sup> *Online Policy Group v. Diebold, Inc.*, 337 F. Supp. 2d 1195, 1204 (N.D. Cal. 2004).



notice.<sup>280</sup> Of course, given the increasing computational power and bandwidth that are accessible and available, one may challenge those assumptions in the first place. Nonetheless, owners and reporters have asserted that notwithstanding, their automated systems “can reliably and accurately identify content without downloading the file itself.”<sup>281</sup> This is of course an untrue assertion because research has shown that a not insubstantial number of the allegedly infringing files linked to by cyberlockers sites are incorrectly categorized or password-protected, making them inaccessible or unverifiable.<sup>282</sup>

That so many reporters have failed the Megaupload test (and yet others have passed the same test) throws this unqualified assertion into question. Even if these could be dismissed as programmatic aberrations or only a reflection of a small number of all takedown requests issued, these observations combined are troubling for a more important reason. They call into question the care or diligence of the reporters and their unqualified reliance on their automated processes. They also call into question the standards applied by online service providers and their diligence in screening and acceding to takedown requests. Given the publicity associated with the incident, the Megaupload test certainly suggests that the reporters actively targeted the disabled sites and other sites that hosted links to these disabled sites and made no distinction between the two. And it is by no means the last. Surely other problems await further discovery. By seriously undermining the “honor system” under which the content and online service provider industries have so far observed, this in turn raises larger questions as to the workings of the DMCA takedown mechanism. With the empirical data at hand, perhaps it is time to revise it for the better.

---

<sup>280</sup> Declaration of David Kaplan in Support of Warner Bros. Entertainment’s Motion for Summary Judgment at 6, *Disney Enter., Inc. v. Hotfile Corp.*, No 1:11-cv-20427-KMW (S.D. Fla. Feb. 28, 2012), ECF No. 308-1; *see also* *Disney Enter., Inc. v. Hotfile Corp.*, No. 11-20427-CIV, 2013 WL 6336286, at 17 (S.D. Fla. Sept. 20, 2013) (noting that Warner admitted that its staff did not download or review any Hotfile content before marking it for removal for computational resource reasons).

<sup>281</sup> Declaration of David Kaplan in Support of Warner Bros. Entertainment’s Motion for Summary Judgment at 6, *Disney Enter., Inc. v. Hotfile Corp.*, No 1:11-cv-20427-KMW (S.D. Fla. Feb. 28, 2012), ECF No. 308-1.

<sup>282</sup> Lauinger et al., *supra* note 213.

## IV. PROPOSALS FOR REFORM

In the 21 years since its enactment, the one component of the DMCA that has really undergone the test of time is the takedown notice mechanism. It is the lynchpin which underlies the delicate balance of responsibilities between the copyright owner and the online service provider by unambiguously placing the primary responsibility for policing the Internet on the copyright owner. In exchange, it has afforded both large content providers as well as individual copyright owners with an accessible and relatively cost-effective<sup>283</sup> means to seek the removal of infringing materials online. If the takedown mechanism is an exercise in individual empowerment and democratization, it has been a resounding success.

One positive change arising from the industry's greater experience with takedown notices has been the increasing use of electronic forms for submitting takedown notices. The displacement of emails in favor of web forms has all but eliminated notices with errors like signatures and attestations.<sup>284</sup> There is considerable room for well-implemented web forms to detect notices that do not comply with the functional formalities such as missing descriptions of the copyright work or missing or inaccurate location information (the URI) for the allegedly infringing material. As noted above, errors in functional formalities amounting to *at least* 9.8% of all notices issued between 2011 and 2015 are easily correctable, as demonstrated by the use of web forms.<sup>285</sup> And indeed, the fact that the Lumen database has ceased to record this information<sup>286</sup> is an indirect testament to the fact that errors with signatures and attestations are no longer an issue because of well-implemented web forms and notice validation mechanisms.

However, one negative impact, which appears to stem from the industry's increasing reliance on automated means for detecting and

---

<sup>283</sup> Note that "cost effective" can be relative; smaller copyright owners such as independent filmmakers and small recording labels may find the effort involved in sending out high volumes of repetitive takedown notices daunting, as well as a distraction from their actual revenue-earning work. Stephen Carlisle, *DMCA "Takedown" Notices: Why "Takedown" Should Become "Take Down and Stay Down" and Why It's Good for Everyone*, NOVA SOUTHEASTERN UNIVERSITY (July 23, 2014), <http://copyright.nova.edu/dmca-takedown-notices/> [hereinafter *Carlisle*].

<sup>284</sup> See *supra* Table 3 and accompanying discussion.

<sup>285</sup> See *supra* Table 3 and accompanying discussion.

<sup>286</sup> See *supra* note 73 and accompanying text.

reporting infringement, is that there are potentially large numbers of notices with substantive errors. Between 2011 and 2015, almost 2.40 million takedown complaints have been sent which have not documented the copyright owner of the infringed work.<sup>287</sup> 4.89 million takedown complaints without takedown requests have been sent.<sup>288</sup> And 2.74 million non-actionable takedown requests have been sent and directed against defunct cyberlocker sites.<sup>289</sup> And these numbers merely represent the *lower bound* of the actual rates of substantive errors, because there could very well be more notices that misidentify the copyright owner, or have no takedown requests, or target defunct sites. After all, the tests used above to illustrate the different types of substantive errors that are possible are certainly not exhaustive.

The use of automated solutions by right holders was clearly designed to improve human productivity in the task of reviewing the countless online resources for possible unlicensed use.<sup>290</sup> It cannot be gainsaid that these takedown systems were designed to specifically reduce human intervention.<sup>291</sup> But, “the full benefit of these technologies will be attained only if they are aligned with our defined values and ethical principles.”<sup>292</sup> Here, the values which we seek to protect are the very values which led to the enactment of the DMCA—to harmonize the promotion and protection of the freedom of opinion and expression and creativity on the Internet as a human right<sup>293</sup> with the legitimate interests of the right holders to protect the creation and use of such intellectual expressions. These goals are mutually supportive.<sup>294</sup> What this study has shown is that the use of automated

---

<sup>287</sup> See *supra* Table 4.

<sup>288</sup> See *supra* Table 4.

<sup>289</sup> See *supra* Table 9; see *supra* Figure 7.

<sup>290</sup> Seng, *supra* note 19, at 432.

<sup>291</sup> *Id.*

<sup>292</sup> INST. OF ELEC. AND ELECS. ENG'RS, ETHICALLY ALIGNED DESIGN: A VISION FOR PRIORITIZING HUMAN WELL-BEING WITH AUTONOMOUS AND INTELLIGENT SYSTEMS 6 (Vol. 2 2018), [https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead\\_v2.pdf](https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf) [hereinafter *Ethically Aligned Design v2*].

<sup>293</sup> U.N. GAOR, 32nd Sess. at 1, U.N. Doc. A/HRC/32/L.20 (June 27, 2016), [https://www.article19.org/data/files/Internet\\_Statement\\_Adopted.pdf](https://www.article19.org/data/files/Internet_Statement_Adopted.pdf).

<sup>294</sup> H.R. REP. NO. 105-551 (II), at 23 (1998) (“The debate on this legislation highlighted two important priorities: promoting the continued growth and development of electronic commerce; and protecting intellectual property rights. These goals are mutually supportive.”).

solutions has been allowed to operate unchecked, and without clear metrics to ensure their transparency and accountability to the Internet community and to mitigate their opportunities for misuse and abuse.

It ought to be noted that these tests above certainly do not set out to review any algorithmic shortcomings of “over-classification”—wrongly targeting sites as hosting infringing content. Instead, the tests focus on the human element—the failure on the part of reporters to check and validate the input information in the notices before their submission. As regards the misidentification of the copyright owner, this appears to be caused by simple human error in inputting the wrong information into the “robo-takedown” system.<sup>295</sup> As regards the Megaupload test, this appears to be caused by reporters who failed to take the additional (and critical) step of verifying that the URIs in their takedown requests actually identify material that is the subject of infringing activity.<sup>296</sup> Such a situation is not helped by the fact that for many reporting agents, the number of issued takedown requests serves as a key benchmark, if not the only benchmark, for their remuneration.<sup>297</sup>

Three simple changes to the DMCA may however suffice to address many of these problems. The first proposal is to make a subtle but important change to the existing language of the DMCA to require a reporter issuing the takedown notice, under penalty of perjury, to attest to the accuracy of the information in the notice and its good faith belief of its claims of copyright infringement. Currently, the DMCA only calls for the reporter's statements of accuracy and good faith belief without making them attestations under penalty of perjury.<sup>298</sup> It is clearly incongruous to only require the reporter to attest to its authority

---

<sup>295</sup> See *supra* notes 224–227 and accompanying text.

<sup>296</sup> See *supra* Table 9; see *supra* Figure 7; see *supra* note 268 and accompanying text.

<sup>297</sup> For example, DMCA.com states that takedowns for up to 25 URLs per website or domain will cost \$199.00 per takedown; presumably, more URLs or domains will increase the cost proportionately. *How Much Will My Takedown Cost*, DMCA.COM, <https://www.dmca.com/faq/How-much-will-my-Takedown-cost> (last visited Sep. 17, 2019).

<sup>298</sup> Cf. 17 U.S.C. § 512(c)(3)(A)(v), (vi) (2010), with 17 U.S.C. § 512(c)(3)(A)(vi) (2010).

to act and hold it to a much lesser responsibility for all other relevant and pertinent information that it has supplied to the service provider.<sup>299</sup>

The second proposal is to provide a mechanism to require reporters to submit verified takedown requests. This will go some way to address not just the problem identified by the Megaupload test, but also alter the existing practice that simply assumes a reliable and accurate identification of the infringing content without accessing or downloading the infringing content itself. Content owners have always complained about the “whack-a-mole” problem wherein a disabled link to allegedly infringing content reappears in a new link.<sup>300</sup> The Megaupload test suggests that this “whack-a-mole” problem may be less intractable than the content owners have suggested, because a not-insignificant number of these links to allegedly infringing content may be non-functional links after all.

Of course, it may be argued that removing these non-functional links harms no one—not the content provider, and certainly not an information location service provider like Google. This is not true. The service provider’s resources that could otherwise be deployed to more carefully process the functional links will be diverted to process the non-functional links. In addition, a reporter that targets sites that host non-functional links is also wasting its own resources to no end. It is only in the copyright owner’s interest to target the functional links; removing non-functional links from a search engine only helps the pirating-end user by making his searches more precise and more likely to be fruitful.

This may be achieved by revising the formal requirement in section 512(c)(3)(A)(iii) of the DMCA to require a reporter to verify the takedown URIs and the dates of verification. In practical terms, this would require the reporter to attempt to access the URIs pertaining to

---

<sup>299</sup> That it is clearly incongruous can be illustrated by the fact that some judges make the mistake of reading the penalty of perjury to apply to both the statement of authorization *and* the statement of good faith belief. *See* Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1112 (9<sup>th</sup> Cir. 2007) (“The DMCA requires a complainant to declare, under penalty of perjury, that he is authorized to represent the copyright holder, and that he has a good-faith belief that the use is infringing.”).

<sup>300</sup> Carlisle, *supra* note 283.

the takedown request.<sup>301</sup> Testing the URI as opposed to actually downloading the resource referred to in the URI is less bandwidth intensive, can be done quickly and efficiently, and is better than the current industry practice of not requiring any download or review of the targeted content before the takedown notice is issued.<sup>302</sup> By requiring reporters to indicate if they have validated the URI as “information reasonably sufficient to permit the service provider to locate the material,”<sup>303</sup> this amendment reasserts the need for contemporaneity as an aspect of accuracy in takedown requests. In most instances, particularly the small copyright authors, they would have taken this obvious step already, and so this would not impose any additional reporting burdens on these author-reporters. In fact, some reporters already provide this information voluntarily,<sup>304</sup> and so turning this into a mandatory requirement levels the playing field for all reporters.

The third proposal calls for a mechanism to place a “cost”—a binding disincentive—on reporters for submitting bad or erroneous takedown requests. In the rush to stem the tide of piracy and in the absence of penalties for making “false positive” takedown requests, reporters have tended to file takedown requests which a judge has described in one case as “overzealous and overreaching.”<sup>305</sup> If the marginal price of each arrow is near zero, to improve his chances, the reporter will fire off as many arrows as he can to hit a target, regardless of the accuracy or precision. But with thousands of reporters doing the same thing, there is a real danger that their actions will either bring a

---

<sup>301</sup> If the URI cannot be accessed, the HTTP will return an error code, or indicate if the URI has been redirected to some other URI. The former shows that the URI is inaccessible, and the latter allows the reporter to target the resource directly. Some highly suspect sites take steps to disable access to their sites when their systems detect that they are being probed by systems owned by owners and reporters. But if this probing step is avoided, there is no way for a reporter to determine if the targeted resource is actually available and therefore actually engaged in infringing activity.

<sup>302</sup> See *Disney Enterprises, Inc. v. Hotfile Corp.*, No. 11-20427-CIV, 2013 WL 6336286, at 15 (S.D. Fla. Sept. 20, 2013).

<sup>303</sup> 17 U.S.C. § 512(c)(3)(A)(iii).

<sup>304</sup> See, e.g., *DMCA (Copyright) Complaint to Google*, CHILLING EFFECTS (Nov. 13, 2012) (on file with author) (reporting by Peer Media Technologies contains time-stamp of when infringing URL was found).

<sup>305</sup> See *Disney Enterprises, Inc. v. Hotfile Corp.*, No. 11-20427-CIV, 2013 WL 6336286, at 16 (S.D. Fla. Sept. 20, 2013).

legitimate but targeted online service provider to its knees,<sup>306</sup> or compromise and cause it to conduct minimal review of these submissions. Courts are beginning to recognize these dangers by ruling that the issuance of numerous defective takedown notices may be grounds for the recipient service provider to mount an action for knowing misrepresentation under section 512(f).<sup>307</sup> Yet the jurisprudence in this area is relatively untested and the service provider may find it difficult to surmount the requirement to prove damages.<sup>308</sup>

A better solution then is to create a two-tier system for handling takedown requests: a normal tier for most takedown requests, and a “slow lane” for handling takedown requests from specific reporters who have been responsible for the most egregiously bad requests. In other words, the penalty for submitting erroneous takedown requests is that they will be given lower priority and handled more slowly than most other requests. Most online service providers already track takedown submissions for each reporter,<sup>309</sup> and so the groundwork has already been laid for the implementation of this solution. As earlier mentioned,<sup>310</sup> Google has in a documented instance withdrawn two reporters from its TCRP.<sup>311</sup> Various criteria can be used to determine when a reporter will be brought within the “slow lane” class of reporters. For instance, the error rates of requests submitted by reporters could be tracked, and the top 10% of reporters with the highest weekly moving average error rates that exceed a stipulated

---

<sup>306</sup> Eric Goldman, *Are Takedowns in 48 Hours “Expeditious” Enough?—Square Ring v. UStream*, TECHNOLOGY & MARKETING LAW BLOG (Jan. 27, 2015), <http://blog.ericgoldman.org/archives/2015/01/are-takedowns-in-48-hours-expeditious-enough-square-ring-v-ustream.htm> (noting that service providers cannot make copyright infringing activity magically disappear at no cost with zero turnaround times).

<sup>307</sup> See, e.g., *Rosen v. Hosting Services, Inc.*, 771 F. Supp. 2d 1219 (C.D. Cal. 2010).

<sup>308</sup> See, e.g., *Lenz v. Universal Music Corp.*, No. 5:07-cv-03783-JF, 2013 WL 271673, at \*9 (N.D. Cal. Jan. 24, 2013) (holding that “any damage” in § 512(f) encompasses damages even if they do not amount to substantial economic damages); *Automattic Inc., v. Steiner*, 82 F. Supp. 3d 1011, 1030 (N.D. Cal. Mar. 2, 2015) (holding that the online service provider entitled to recover damages for time and resources incurred in dealing with the defective takedown notices, in the form of employees’ lost time and attorneys’ fees).

<sup>309</sup> See, e.g., *Content Delistings Due to Copyright*, GOOGLE, <https://transparencyreport.google.com/copyright/overview?hl=en>.

<sup>310</sup> *How Google Fights Piracy*, *supra* note 25, at 17.

<sup>311</sup> *Id.*

error rate of, say, 5%, will fall into the “slow lane.” Such a system will not discriminate against the small copyright owner or first-time reporter, because the moving average error rate will be assessed over, say, a substantive period of 6 months. Nor is the system necessarily stacked against the largest users of the takedown system by volume of takedown requests, because it is possible to issue large volumes of takedown requests and have low error rates, as the empirical studies above have shown. And to ensure that the criteria will not be used by service providers to serve as a shield from processing takedown requests, it could also be provided that the two-tier system could only be introduced by online service providers who can prove that they receive more than a certain threshold by volume of takedown notices.

Admittedly, this two-tier system with its “accountability metrics”<sup>312</sup> will not be popular with content providers. But introducing a benchmark setting measure will help ensure better verified notices, lower the incidents of “false positive” takedown requests, promote accountability of reporters to right holders and to service providers, raise the overall standards for reporters and encourage a “race to the top.” This will in turn reduce the incidents of indiscriminate and “whack-a-mole” takedowns and encourage the development of more precise technologies for detecting and reporting online infringements which are less likely to generate false positives, and are more measured and nuanced when compared to the takedown-and-stay-down mechanism. After all, the use of value-based design methodologies is a hallmark of good engineering, as it attempts to capture and codify societal well-being, social costs and overall economic value into metrics which can be empirically evaluated by oversight systems for compliance.<sup>313</sup> And such methodologies could even serve as the prelude for an industry wide technical solution which can definitively address the problems of the proliferation of unlicensed works, a solution which has hitherto not been explored though it has been embedded in the DMCA.<sup>314</sup>

---

<sup>312</sup> *Ethically Aligned Design v2*, *supra* note 292, at 8.

<sup>313</sup> *Id.* at 55.

<sup>314</sup> 17 U.S.C. § 512(i)(1)(B) (2010) (providing that online service providers are to accommodate and not interfere with standard technical measures).



## CONCLUSION

If a history of the DMCA is ever to be written, one of its most defining moments will be when copyright owners and reporters started to use automated solutions to detect instances of copyright infringement and submit takedown notices and their requests to the service providers. This is perfectly understandable. As online piracy proliferates, we have to use automated solutions to find, manage and contain this serious problem. Conversely, the opportunities for human intercession have been reduced. But this is not to say that human input is less important. On the contrary, human input is vital because it shapes the way we design our automated systems and the results we want from them.

Unfortunately, for takedown notices this interaction between human review and automated processing has broken down. It has created a situation where erroneous notices that fail to comply with non-functional formalities like electronic signatures and attestations are (almost) non-existent, because these procedural elements can be purely automated. But where there is ostensibly human input, there are a substantial number of notices (5.5%) between 2011 and 2015 that fail to properly comply with the functional formalities such as having a description of the copyrighted work.<sup>315</sup> Likewise, the significant number of notices with empty takedown requests (9.8%) or a smaller but not insignificant number of improperly validated takedown requests (0.23%) again show a systemic failure on the part of the reporters to properly configure their takedown programs.<sup>316</sup>

We may condone some mistakes, particularly mistakes which are innocent and reasonable. But we ought to impose penalties for mistakes and escalate them into liabilities against the reporting agents where these mistakes were caused by the unreasonable behavior of the reporters.<sup>317</sup> These errors were never caught for months on end. We cannot, and should not, allow parties to elude responsibility, just because they use machines to extend their decision making

---

<sup>315</sup> See *supra* Table 4.

<sup>316</sup> See *supra* Table 4.

<sup>317</sup> See *Disney Enterprises, Inc. v. Hotfile Corp.*, No. 11-20427-CIV, 2013 WL 6336286, at 47–48 (S.D. Fla. Sept. 20, 2013) (holding reporters potentially liable for misrepresentation under § 512(f) for egregious and unreasonable mistakes made in intentionally taking down content on defendant cyberlocker site for which they did not own the copyright).

capabilities.<sup>318</sup> For this reason, we have to expose as completely facetious the argument that the reporter is not responsible for its computers as “computers conducting automated searches cannot form a belief consistent with the language of the DMCA, because they cannot distinguish between infringing content and content that merely contains words that suggest infringement.”<sup>319</sup> In truth, it is the reporters who are responsible—the same reporters who, in the quest for speed and efficiency, have arrogated their individual human responsibilities to the programmers and their codes and scripts.<sup>320</sup> They have, through the largely automated mechanism of the DMCA, enabled the illicit enforcement of copyrights—a “copywrong.”<sup>321</sup> Surely this calls for greater scrutiny of their takedown processes, and not less.

In today’s increasingly automated world, although machines have taken over much of the mundane and repetitive tasks to which we

---

<sup>318</sup> See, e.g., *Rossi v. Motion Picture Ass’n of Am. Inc.*, 391 F.3d 1000, 1005 n.7 (9th Cir. 2004) (noting that even though MPAA used the “Ranger” program to identify potentially infringing websites, it was the employees of MPAA who ultimately decide whether a website contained infringing material).

<sup>319</sup> *Id.*

<sup>320</sup> For an example of the discussion regarding imputing liability to manufacturers of autonomous technologies, see the discussion in Nevada, the first state in the U.S. to adopt regulations regarding the use of autonomous car manufacturers, regarding liability issues for accidents caused by autonomous technologies. INFO. STATEMENT OF ADOPTED REG. AS REQUIRED BY ADMIN. PROCEDURES ACT NRS 223B.066 LCB FILE NO. R084-11, at 31 (Feb. 6, 2012), <http://www.leg.state.nv.us/register/2011Register/R084-11A.pdf> (taking the position that autonomous vehicle manufacturers will be held liable for defects and non-compliance issues under the law of product liability); see John Markoff, *Collision in the Making Between Self-Driving Cars and How the World Works*, N.Y. TIMES (Jan. 23, 2012), <http://www.nytimes.com/2012/01/24/technology/googles-autonomous-vehicles-draw-skepticism-at-legal-symposium.html>; see also JOHN VILLASENOR, BROOKINGS INSTITUTE, PRODUCTS LIABILITY AND DRIVERLESS CARS: ISSUES AND GUIDING PRINCIPLES FOR LEGISLATION, (2014), [https://web.archive.org/web/20160310063344/http://www.brookings.edu/~media/research/files/papers/2014/04/products%20liability%20driverless%20cars%20villasenor/products\\_liability\\_and\\_driverless\\_cars.pdf](https://web.archive.org/web/20160310063344/http://www.brookings.edu/~media/research/files/papers/2014/04/products%20liability%20driverless%20cars%20villasenor/products_liability_and_driverless_cars.pdf).

<sup>321</sup> The term originated as the title of a 1993 article by Richard Stallman, decrying the taxation on equipment permitting the making of home copies of music; his argument was that such taxation in the name of copyright was not justified and therefore became a “copywrong.” See Richard M. Stallman, *Copywrong*, WIRED (Mar. 1 1993), <https://www.wired.com/1993/03/1-3-stallman-copyright/>.

used to apply manual labor, there is still much we can do. We commend the Internet intermediaries and the operators of the then-Chilling Effects—and now Lumen—repository, who believe that transparency is one of the best ways to ensure accountability for online decisions. Let us build upon their efforts, systematically evaluate and review these takedown notices, put a stop to these errors and hold the reporters to a higher standard. Let us translate the lessons we have learnt from this study to develop an improved takedown system: one that promotes openness and transparency, encourages the development of standards, invites competitive innovations and demonstrates better accountability. Let us end this process of copyrighting copywrongs.