

# Beyond the DMCA: How Google Leverages Notice and Takedown at Scale

By Caleb Donaldson

Caleb Donaldson is copyright counsel at Google.

As the Internet continues to grow into every corner of modern life, we face a challenge. More people are making more writing, imagery, and music than ever before, and sharing it more widely. Yet at the same time, these very same digital technologies allow perfect infringing reproductions of copyrighted materials, posing a challenge to many who would try to make a living creating content people love. Online intermediaries such as Google have a role to play. Billions of users rely on us to help them locate information in the ever-growing datasphere. The Digital Millennium Copyright Act (DMCA), a 1998 compromise between a desire to foster the growth of the Internet and a need to protect the interests of copyright holders, is a cornerstone of Google's ability to continue to help make the world's information universally accessible and useful. In so doing, we can also help those who seek to create and share new works with the world. This article lays out the basics of the DMCA's notice-and-takedown regime, discusses the challenges of doing removals at scale, and explains some of the ways Google uses notice and takedown that go beyond what the law requires to help fight piracy.

The scale of Google's Web Search operation today is vast. Google crawls more than 20 billion web pages every day. We have found more than 130 trillion web addresses so far, and we have indexed pages from more than 165 million different domains. We perform trillions of searches for our users every year, providing useful links to the farthest corners of the sprawling web in an average of about a quarter of a second. The vast majority of these links lead to perfectly legitimate content, of course, but there are some bad apples in the barrel. To identify these, we need rights holders' help.

## The DMCA

The DMCA was enacted in 1998, the same year Larry Page and Sergey Brin founded Google in Menlo Park. Online services were beginning to show tremendous promise. But the nature of copyright law and the statutory damages regime meant that even though most content was legitimate, hosting or linking to even a small amount of infringing material could create crippling liability for intermediaries. The goal of the DMCA was to provide a way for copyright holders to protect their works online, while also providing the legal certainty necessary for continued investment in, and growth of, the Internet.<sup>1</sup> The safe harbor provisions in the DMCA embody the essence of this compromise.<sup>2</sup> On the one hand, rights holders can achieve "expeditious" removal of allegedly infringing content, without the need to register the works with the Copyright Office or file a lawsuit—without even involving a lawyer to write a cease and desist letter. On the other hand, online service providers, as defined in the DMCA, do not have to expose themselves to the risk of company-ending liability, so long as they remove allegedly infringing content

upon notice and adhere to the other requirements of the safe harbors. This notice-and-takedown scheme makes sense because rights holders are in the best position to know whether their content is being used without authorization. And this bargain made the modern Internet possible. Search engines could direct users to relevant web pages, and user-generated content platforms could publish the materials their users uploaded, all without fear of crushing statutory damages awards.<sup>3</sup> In this working model, Google Web Search has thrived.

### **Google Web Search**

Currently, Google Web Search receives DMCA complaints for three million or so URLs a day. That may seem like a big number in isolation, but remember the scale of Google Web Search as a whole. Thanks in part to initiatives such as our high-volume submission program for trusted submitters, the Trusted Copyright Removal Program (TCRP), we are able to keep pace with DMCA complaints even as the web and our Web Search index grow. Google also created the TCRP as a recognition that different kinds of submitters had different needs. Smaller-scale submitters tended to commit proportionately more errors and valued an easy, accessible way to submit requests. Larger-scale submitters valued tools to submit URLs at high volumes, and required different kinds of assistance and oversight.

While the vast majority of copyright takedown notices that we receive are well founded, there are a small number that are not. Google invests significant resources to process the well-founded notices promptly, and also to identify those that are erroneous or abusive. Google will push back if we suspect a notice is mistaken, fraudulent, or abusive, or if we think fair use or another defense excuses that particular use of copyrighted content. Some examples of mistaken and abusive takedowns include:

- An individual claiming to be a candidate for political office in Egypt filed a copyright complaint to delist two pages on Egyptian news sites reporting on the individual's arrest record.
- An antipiracy enforcement firm representing a music label filed a copyright complaint asking us to delist dozens of web pages containing the word "coffee" in the title. These URLs had nothing to do with the identified copyrighted work.
- A Ukrainian politician sent us a copyright complaint regarding the use of his image in a number of articles critical of his performance in office.
- A business filed a copyright complaint regarding the use of images of its products in a Techdirt article discussing that business's attempt to make Google delist a number of pages.
- A poet sent repeated takedown notices targeting criticism and commentary relating to the poet's online copyright enforcement efforts.
- A well-known publisher of children's books sent a takedown notice targeting the use of excerpts by a critic discussing the use of gun imagery in children's literature.
- A physician claiming a copyright in his signature sent a takedown notice aimed at a document related to the suspension of his license to practice medicine.

Nonetheless, we approve more than 97 percent of the Web Search complaints we receive, and our average time to resolution is still under six hours. In fact, the vast bulk of the notices we receive comes from a small group of trusted partners, typically rights holders' industry associations (like the Recording Industry Association of America and the Motion Picture Association of America here in the United States, and their foreign counterparts) or third-party businesses that provide antipiracy services. These trusted partners have helped us to refine our processes and policies, even as they have scaled up their own submissions dramatically.

### **Leveraging Notice and Takedown**

The DMCA provided Google and other online service providers the legal certainty they needed to grow. And the DMCA's takedown notices help us fight piracy in other ways as well. Indeed, the Web Search notice-and-takedown process provides the cornerstone of Google's fight against piracy.

### **Demotion Signal**

Google counts the number of URLs in a given domain that have been removed pursuant to takedown notice complaints. As that number increases, we use it as a signal in the ranking of the entire domain in Web Search, meaning that the whole domain is less likely to figure prominently in our results. This "demotion signal" means that even URLs which have never been the subject of notices can be affected by the notice-and-takedown process. This reduces the burden on rights holders to keep sending notices for rogue sites. As mentioned above, it is the rights holders who are in the best position both to know whether a particular URL infringes their copyrights, and to decide which domains to target for enforcement. The demotion signal is a way to amplify the power of every notice, with each submitted URL having a penumbral effect on its entire domain.

### **Ads Policy Enforcement**

Google firmly believes that a "follow the money" approach must be a central part of fighting piracy. Large-scale, commercial pirates will be motivated to circumvent other obstacles, and will engage in an endless game of cat and mouse until we can remove their financial incentives. The Police Intellectual Property Crime Unit in the United Kingdom has estimated that shutting off advertising revenue would close 95 percent of these infringing sites.<sup>4</sup> To further this effort, Google makes use of the takedown notices rights holders send us in several ways.

When a URL is removed from Search results, that URL is automatically disqualified from carrying advertisements from our AdSense network. This makes it harder for web publishers to monetize infringing pages. And our ads policy enforcement team looks at domains with large numbers of Web Search DMCA complaints for policy compliance. Overall, since 2012, we have blacklisted more than 91,000 sites from our AdSense program for violations of our copyright policy. We have also terminated more than 11,000 AdSense accounts for copyright violations. Thus, every DMCA notice is made a little more powerful still, and the vast stream of notices we process becomes a source of data for our other antipiracy enforcement efforts.

Google similarly prevents any AdWords ads (the ads that appear on Search results pages), from linking to URLs we have removed from Search results pursuant to a DMCA notice. This means an infringing publisher cannot pay Google for an ad that sends traffic to a page we have removed from Search results for copyright infringement.

### **Transparency**

We believe that our users should know when and why we remove content from Search results, and the stream of DMCA notices offers valuable insights. We forward Web Search DMCA notices to Lumen, a third-party nonprofit research project of the Berkman Klein Center for Internet & Society at Harvard University, in conjunction with other partners. Lumen redacts personally identifying information and publishes the notices. Researchers and academics have made use of this data to understand the state of the web and the continued viability of the notice-and-takedown regime. Additionally, Google publishes much of the same information in the Google Transparency Report.<sup>5</sup> Users can explore the data on the web, or download the entire data set for their own research.

One recent study of the Lumen database looked at a six-month period, analyzing a random sample from the more than 100 million notices reflected in the database for that time period: 99.8 percent of those notices were to Google Web Search.<sup>6</sup> The study provides valuable insight about the type of organizations sending notices, which creative industries those organizations represent, what kinds of sites are targeted, and the rates of various kinds of deficiencies in the notices themselves. This kind of research is an invaluable aid, made possible by our commitment to transparency and the hard work of Lumen and the researchers.

### **Not-in-Index URLs**

Google has critically expanded notice and takedown in another important way: We accept notices for URLs that are not even in our index in the first place. That way, we can collect information even about pages and domains we have not yet crawled. We process these URLs as we do the others. Once one of these not-in-index URLs is approved for takedown, we prophylactically block it from appearing in our Search results, and we take all the additional deterrent measures listed above. We recently discovered that some bulk submitters make very heavy use of this feature. In one sample we found that around 82 percent of the URLs we approved were not in our index (and have therefore never appeared in any search results).<sup>7</sup> How this discovery will influence the further evolution of our processes, only time will tell. It does suggest that the number of takedown notices we get is not a good proxy for the number of allegedly infringing links we serve.

### **Moving Forward**

Notice and takedown plays a vital role in the battle against online piracy. It is the bedrock layer of a complex topology laid down through years of practice and has spawned its own industry of third-party notice senders. This enables even less sophisticated rights holders to avail themselves of notice and takedown. Google has leveraged the notices it receives to go beyond what the law ever required. As we move into a world where artificial intelligence can learn from vast troves of data like these, we will only get better at using the information to better fight against piracy.

The passage of the DMCA marked a turning point in the development of the Internet in the United States. The law, in describing the various safe harbors, marked out particular paths for emerging digital services to follow, influencing which particular services would flourish. In the almost 20 years since the notice-and-takedown process was codified, the Internet has grown almost unimaginably, both in size and in its variety, offering new services and new ways of connecting people. There is inarguably more content—more text and music and video and imagery—being created and shared around the world than ever before. Much of that is due to the central bargain struck in the DMCA that has fostered collaboration: Rights holders will tell online service providers where their content is infringed, and online service providers will expeditiously remove content at those locations. And just as new and wonderful services and products have been built on top of the basic communications protocols of the Internet, Google has built on the basic notice-and-takedown scheme, leveraging the notices we receive to fight piracy more effectively across our products.

### Endnotes

1. S. REP. NO. 105-190, at 1 (1998) (stating that the purpose of the DMCA is “to facilitate the robust development and world-wide expansion of electronic commerce, communications, research, development, and education in the digital age”); H.R. REP. NO. 105-551(II), at 49 (1998) (observing that the DMCA “preserves strong incentives for service providers and copyright owners to cooperate to detect and deal with copyright infringements that take place in the digital networked environment”).
2. 17 U.S.C. §§ 512 *et seq.*
3. In the United States, statutory damages for copyright infringement range from \$750 to \$30,000 per work infringed in normal cases. 17 U.S.C. § 504(c). In a digital world, where reproduction and linking are practically costless, statutory damages can quickly mount to forbidding heights.
4. Mike Weatherley, *Strides in the Right Direction*, TRADEMARKS & BRANDS ONLINE (June 17, 2015), <http://www.trademarksandbrandsonline.com/article/strides-in-the-right-direction>.
5. GOOGLE TRANSPARENCY REP., <https://www.google.com/transparencyreport/removals/copyright/> (last visited Sept. 21, 2017).
6. Professors Urban, Karaganis, and Schofield published a paper comprising three separate studies about notice and takedown. The second study focuses primarily on DMCA notices sent to Google Web Search. Jennifer M. Urban, Joe Karaganis, & Brianna L. Schofield, *Notice and Takedown in Everyday Practice* 77–96 (UC Berkeley Public Law Research Paper No. 2755628, 2017), <https://ssrn.com/abstract=2755628>. This paper not only has informed Google’s own decisions, but also has been cited in Copyright Office proceedings regarding the DMCA. *See Section 512 Study*, COPYRIGHT.GOV, <https://www.copyright.gov/policy/section512/> (last visited Sept. 21, 2017).
7. *See* Additional Comments [Amended] at 7, Section 512 Study, Docket No. COLC-2015-0013 (Google Feb. 21, 2017), <https://www.regulations.gov/document?D=COLC-2015-0013-92487>.

## **Submitting Removal Notices to Google**

If you believe something on Google's services violates your copyright, you can file a takedown notice by accessing the Google Legal Troubleshooter (<https://support.google.com/legal/troubleshooter/1114905?hl=en>), selecting the product at issue, and following the prompts.

### **Tips for successful submissions:**

1. Provide a description of (or even better, a link to) the allegedly infringed work.
2. Provide the specific URL you wish removed from Google Search results (for a takedown notice about Web Search) or the specific URL where the allegedly infringing content resides on Google's services.
3. Provide a real person's name and electronic signature.