News Corp

November 26, 2021

Comments of News Corporation
to the Copyright Office

Re: Publishers' Protections Study

News Corporation ("News Corp") appreciates the opportunity to submit these first-round comments in response to the Notice and Request for Public Comment issued by the Copyright Office on October 12, 2021 (the "Notice").[1] At the request of Congress, the Copyright Office has initiated a "Publishers' Protections Study" to consider "the effectiveness of publishers' existing rights in news content" both under federal copyright law and other federal and state laws, the desirability and scope of additional protections, and how such new provisions would interact with existing law and treaty obligations.

News Corp is a global diversified media and information services company comprising businesses across a range of media, including news and information services, digital real estate services, book publishing, and subscription video services. Our news publishing businesses in the United States include, among others, the leading daily newspaper *The Wall Street Journal,* published since 1889,[2] and the *New York Post*, the oldest continuously published news publication in the United States, founded in 1801 by Alexander Hamilton.[3] Like all publishers of news content, News Corp depends on intellectual property laws to maintain, support, and protect its endeavors.

As a matter of sound public policy, and to meet the original objectives and intent of the Copyright Clause in the Constitution, the laws must reward journalistic effort to incentivize the professional publication of reliable news content for the benefit of the public. But the news media today is in crisis because intellectual property and other laws in the United States have failed to keep up with technological changes brought about by the internet and digital distribution of news and information. This is an existential issue for news publishers, who are beset by declining revenues, loss of journalistic employees, bankruptcies, and resulting "news deserts" largely caused by the disruptive effects of the free-riding republication of news on the internet.[4]

[1] The Office has indicated that a second round of comments will be accepted through January 5, 2022. *See* 86 Fed. Reg. 62215 (Nov. 9, 2021).

[2] Published by News Corp subsidiary Dow Jones.

[3] Published by News Corp subsidiary NYP Holdings.

[4] The University of North Carolina School of Media and Journalism reported that by 2020, 171 U.S. counties had no local newspaper, and nearly half were covered by one newspaper, often a weekly. *See https://www.usnewsdeserts.com/.*

If we are to preserve a robust independent press, the time for action is now to strengthen the incentives that the law provides to produce quality news content. The Publishers' Protections Study is urgently needed and there has never been a more vital time for the Copyright Office to take stock of the effectiveness of news publishers' existing rights in news content. Much of the law creating protection and thus economic incentives for the press is found in the copyright laws, and with appropriate modifications that should continue to be the case. But legislators and policymakers should also think outside the traditional copyright box to take care that changes in technology do not deprive the press of the tools needed to survive. This may require not only changes to Copyright Office procedure and amendments to Title 17, but also updates to the antitrust laws and non-copyright federal laws enacted under other sources of authority that can aid in preserving the independence and viability of the press.[5] Some long-standing fundamentals of copyright law—such as the registration requirement and the definitions of copyrightable content and infringement—will need to be reassessed with the attitude that nothing is sacred.

This comment will first retrace the value contributed to society by a vibrant, professional press conducting independent journalism. It will then outline some of the gaps of principal concern in legal protection. News Corp respectfully reserves the right to comment further and in greater detail as the Publishers' Protections Study proceeds, particularly about potential reforms to modernize the law and address the gaps in protection—a subject on which much work remains to be done by all stakeholders. Accordingly, this round-one comment is directed at topic (i) in the Notice ("The effectiveness of current protections for press publishers under U.S. law").

1. The societal cost of failing to support journalism through appropriate intellectual property laws

It is no accident that the press is the only industry singled out by name in the Constitution for protection (U.S. Const., Amendment 1). As the framers recognized, reporting the news plays a critical role in democratic society. The First Amendment "rests on the assumption that the widest possible dissemination of information from diverse and antagonistic sources is essential to the welfare of the public." *Associated Press v. United States,* 326 U.S. 1, 20 (1945). A free press is an indispensable "condition of a free society." *Id.* The press disseminates reliable information necessary to maintain an informed citizenry and government. It holds power to account by undertaking investigations and performing analysis, presents a diversity of viewpoints, and serves as an incubator of new viewpoints and a voice of change.[6] Investigative journalism is often the catalyst of high-impact criminal, regulatory, and legislative proceedings.

[5] While re-establishing the ability of news publishers to obtain fair remuneration for the use of their content will likely require a re-tuning of multiple legislative frameworks, including competition law and section 230, the work must start with strengthening the protections of copyright law. To be in a position to fairly and effectively negotiate payment for use of their content, news publishers need, first and foremost, a strong property right in that content forming the backbone of their negotiating position.

[6] For example, a recent study by Gao et al., *Financing Dies in Darkness? The Impact of Newspaper Closures on Public Finance*, 135 Journal of Financial Economics 445 (February 2020), examined how local newspaper closures affect public finance outcomes for local governments. Following a newspaper closure, municipal borrowing costs increased by 5 to 11 basis points, costing the municipality an

Professional news organizations are judged by the accuracy, relevance, and fairness of what they publish. We employ trained editors to maintain quality content while upholding ethical codes and complying with the law. We accept responsibility for the material we publish, receive complaints, and correct errors. We are subject to the laws of libel, invasion of privacy, and the like, and have real assets exposed to risk if legal standards are violated—which prompts meaningful, effective governance. Unlike others who can avail themselves of section 230 of the Communications Decency Act, news publishers cannot amplify false and damaging content with impunity.

Gathering and reporting on the news impose heavy costs. News publishers employ journalists, subject-matter experts, photographers and videographers, fact-checkers, editors, and administrative support staff. They incur the cost of maintaining equipment and offices, including in many cases a far-flung global network of correspondents and news bureaus necessary to provide comprehensive coverage of newsworthy events from around the world. Often, a single groundbreaking investigative report may be the result of months or more of time-intensive behind-the-scenes work by a team of reporters.

Those costs are not just financial. Intrepid journalists put their lives on the line every day to collect and report the news from the front lines of conflict. The plight of reporters desperately seeking to exit Afghanistan as it fell to the Taliban earlier this year, and of staff members left behind, is only the latest in a long list of journalistic casualties. Regina Martinez, Gumao Perez and many other reporters slain in Mexico in the last few years, *The Sunday Times'*[7] Marie Colvin (killed in Syria in 2012), and *The Wall Street Journal's* Daniel Pearl (executed by terrorists in Pakistan in 2002) are just some of those who lost their lives reporting the news. Even when death does not result, physical violence against reporters is on the rise: The Reporters' Committee for Freedom of the Press found that the 438 physical assaults on journalists in 2020 was over three times the number it counted for the prior three years combined. *See* https://www.rcfp.org/wp-content/uploads/2021/05/Press-Freedom-Tracker-2020_FINAL.pdf at 8. Aside from physical violence, arrests, detention and prosecution of journalists is a never-ending problem. *See id.* at 12-13.

While news content is expensive to originate, once created the marginal cost of reproducing news reports in copies and electronic displays is dramatically lower. Professional news organizations that produce an intangible product have a limited time to recoup their costs and generate profits, and are particularly at risk from the continuous and systematic copying of the fruits of their labors. Indeed, in the current age, the time news organizations have to charge a premium price based on the exclusivity that results from breaking a story is approaching zero.

The news media are dependent on intellectual property protection for their very existence. As the court stated in *Associated Press v. Meltwater U.S. Holdings, Inc.*, 931 F. Supp. 2d 537, 553 (S.D.N.Y. 2013), "[i]nvestigating and writing about newsworthy events occurring

---

additional $650,000 per issue. According to the authors, "[t]his effect is causal and not driven by underlying economic conditions."

[7] Published in the United Kingdom by News Corp subsidiary Times Newspapers Ltd.

around the globe is an expensive undertaking and enforcement of the copyright laws permits [media organizations] to earn the revenue that underwrites that work."

2. New challenges of the digital age

The relative limitations of pre-internet technology created dynamics that allowed professional news organizations adequate time to profit from their investments and efforts. Slower means of copying and disseminating gave news publishers a natural, valuable period of exclusivity. The limited bandwidth of traditional mass communication vehicles—such as the printed newspaper and licensed television and radio stations—afforded copyists few opportunities to profit from free-riding. Accordingly, a news publishing organization that built a reputation for reliability and timeliness could effectively sell a product that differentiated itself consistently and compellingly by providing these benefits to readers.

With the advent of communication through the internet, the technological dynamics that once protected professional news organizations' limited exclusivity have been overwhelmed. The internet has enabled unprecedented speed and breadth of news dissemination, and copies can be made and transmitted costlessly with a few keystrokes. This has destroyed the pre-existing equilibrium that, with a relatively modest assist from copyright law, enabled publishers to control access to their output and generate compensation for the costs of conducting journalism.

Technological advances easily enable misuse and misappropriation of electronically published news content, permitting free-riders to profit from the hard work and investment of professional news organizations. Consumers today increasingly access and consume news through a wide variety of online intermediaries that reproduce, summarize, and otherwise redisseminate original news content, often in combination from a number of sources.

Online redistributors of original news content pose a real and unprecedented threat to traditional journalism in several ways. They intervene between news publishers and their readers, diverting traffic from the news publisher's properties, enabling the redistributor to obtain the benefits of a first-party relationship with the audience—both to harvest and arrogate to its own purposes valuable user data and to sell advertising for its own benefit. Redistributors also atomize news content into individual items of short-form journalism compiled from multiple sources, so that many consumers' experience is no longer with a "News Publication" as an integrated whole. This results in severe business implications for publishers of original news content, who suffer flattening of their brand and are disincentivized from differentiating their products because many readers who obtain the news on the internet are not even aware of its source. This also has a deleterious effect on readers, because it obscures quality signals they rely upon to differentiate between reliable and unreliable sources of news.

These technological transformations in how news content is delivered have created an ecosystem with perverse economic incentives that reward republishers who do not invest in reporting the news but have positioned themselves to reap the profits from doing so, at the expense of those who do foot the bill for original journalism and are accountable for their reporting. This is an unprecedented threat to the survival of professional news organizations, resulting in mass layoffs and shutdowns of news organizations, a sharp decline in local news outlets, dilution of quality signals and brands, a decline in production of long-form and

investigative journalism, an increase of unreliable news content, muddying of the line between news and opinion, and the general promotion of news content based on its virality rather than its verity or value to the public. If everyone is free to be a republisher, there will be no publishers and nothing left to republish. In this era of "fake news," it is critical to protect credible and trusted journalism. It is more important now than ever that consumers know and be able to trust the source of their news content.

Absent adequate legal protection for the fruits of journalistic enterprise, technology easily allows for misuse and misappropriation, permitting free-riders to profit on the intellectual effort of others without accountability. At the same time, other features of copyright law that may have caused little concern in the pre-internet age have posed increasing obstacles for news publishers seeking to prosper in a digital environment. In short, now that technology is doing less to protect news publishers, copyright law needs to do more. In this context, we turn to discuss certain areas in which the law does not effectively recognize or protect intellectual property rights of news publishers. As will be seen, the current legal regime protects news publishers in theory but often not in practice.

3. Deficiencies in the current law

A. Scope of copyright protection for news content

A great threat to the viability of original news reporting is the continuing ability and evolving practices of online operators to republish news content, sometimes in condensed "snippet" or headline form and in an array of reports from different publishers, without the apparent threat of copyright infringement liability. By scraping and republishing content in real time, these redistributors make themselves first to market alongside the news publisher that breaks the story, who no longer enjoys any period of exclusivity. And by scraping and republishing content from multiple news sources, redistributors effectively make themselves first to market on all stories, presenting to the public a product against which no individual publisher could possibly compete. This phenomenon enables online redistributors to monetize directly and indirectly the commercial value inherent in news content without incurring the costs of reporting the news, and without even paying the original publisher for the right to do so.

News articles are literary works under 17 U.S.C. § 102(a)(1). Like other fact-based works, they are subject to copyright protection even though copyright does not protect facts as such but only their expression, description, selection, and arrangement. As the Second Circuit observed, "[t]hose who report the news undoubtedly create factual works. It cannot seriously be argued that, for that reason, others may freely copy and re-disseminate news reports." *Authors Guild v. Google, Inc*., 804 F.3d 202, 220 (2d Cir. 2015) ("*Google Books*").

Yet, copyright in fact-based works is also regularly described as "thin," *e.g.*, *Feist Publ. Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 349-50 (1991); *Smart Inventions Inc. v. Allied Commun. Corp.*, 94 F. Supp. 2d 1060, 1066 (C.D. Cal. 2000) ("[T]he scope of protection afforded factual works is very thin."), suggesting that users can appropriate with impunity under the copyright laws more of the original than would be the case with fictional literary works. Courts therefore struggle to find the proper dividing lines between infringement, fair use, and non-infringing appropriation of news content. The result is that U.S. copyright law risks

undervaluing the creative elements of news reporting, which include not only the composition of words used to express the facts reported, but the editorial functions of selecting what to report and not to report, which facts are pertinent, which sources of information are authoritative and reliable, what kind of balance to strike between presentation of conflicting viewpoints, which news items take precedence over others in importance, how deeply to go into a story, whether to credit information provided by sources, whether to name those sources, and all the other myriad editorial judgments that shape the report. These are intellectual, creative functions that qualify as "authorship" in the constitutional sense, not just the "sweat of the brow" which alone does not qualify for copyright protection. But these contributions are too easily ignored in a superficial analysis of whether a copyist has taken "facts" or "expression

Nihon Keizai Shimbun v. Comline Bus. Data, Inc., 166 F.3d 65 (2d Cir. 1999), a case that involved pre-internet technology but was decided at the dawn of the internet age, illustrates some of the difficulties of applying standard copyright doctrine to news content. An investment newsletter reproduced portions of 22 news articles taken without permission from the plaintiff's newswire service. The Second Circuit affirmed a finding of infringement as to 20 of the 22, but held that one instance in which the defendant merely copied facts, not expression, was non-infringing, and that in a second instance in which defendant had copied the first paragraph (i.e., the lede) of a six-paragraph article, containing 20% of the verbiage of the original, the copying was not sufficiently substantial to amount to infringement. The Court cautioned that it did not intend to establish a bright-line rule as to the percentage of a news article that could be taken with impunity, but it was inevitable that the 20% threshold has gained some currency as a rule of thumb.

More recently, technology has made possible the ingestion of vast amounts of published content for the purpose of offering searchable databases to users seeking to locate particular content in the ocean of reported news. *Fox News Network, LLC v. TVEyes, Inc*, 883 F.3d 169 (2d Cir. 2018), dealing with video news reports, and *Associated Press v. Meltwater U.S. Holdings Inc.*, 931 F. Supp. 2d 537 (S.D.N.Y. 2013), dealing with textual news content, are generally and correctly regarded as wins for the news media, but leave important questions subject to debate. *TVEyes* involved a copyist that ingested substantially all broadcast news, which enabled its subscribers both to locate specific items through word searches and then to watch, download, and forward by email specific clips of up to ten minutes. Reversing the district court, the court of appeals held that the so-called "Watch" function was not a fair use (even though it was somewhat "transformative" under the first fair use factor in 17 U.S.C. § 107), based on evidence that the Watch function usurped an existing licensable market for news clips and was therefore unfair under the fourth factor.

Meltwater involved an online clipping service that systematically disseminated to its subscribers excerpts from AP newswire articles that met word-search criteria supplied by its subscribers. The excerpts included a verbatim copy of the headline and up to 300 characters of the lede, a hyperlink to the source, a credit of the source, and an additional excerpt of up to 140 characters showing a search term in context. 931 F. Supp. 2d at 545. The content distributed by Meltwater consisted of between 4.5% and 61% of the original word count. *Id*. at 546. The evidence showed that Meltwater subscribers clicked the hyperlinks provided by Meltwater to

access the source articles at a de minimis rate of .08%. *Id*. at 547. On these facts, the court held that Meltwater's activities were not fair use.[8]

Both *TVEyes* and *Meltwater* relied on the fact that the excerpts that the defendant made available to its subscribers were longer than the snippets returned by a typical search engine and therefore more likely to satisfy the user's demand for the original content. *TVEyes* distinguished the *Google Books* decision on the ground that the clips made available in *TVEyes* were much longer than the snippets of books made available by Google, and, "given the brevity of the average news segment on a particular topic[,] likely provide TVEyes's users with all of the Fox news programming that they seek . . . ." 883 F.3d at 179. Similarly, *Meltwater* rejected the defendant's argument that it functioned as a search engine and that the clips it circulated were just like "snippets" that accompany a listing of results in response to a more conventional internet search. 931 F. Supp. 2d at 555, 558-59.

It remains to be seen how courts would decide whether a redistributor that satisfied a consumer's appetite for the daily news with an array of snippets shorter than those offered by Meltwater, or a video clip shorter than that offered by TVEyes, was liable for infringement, and whether courts will continue to apply the quantitative substantiality test of *Nihon.* The suggestion by the majority in *TVEyes* that the defendant's Watch function was "somewhat transformative" because it improved the efficiency with which users could access content will no doubt be seized upon by those who offer technical advances in content delivery as a plus factor in the fair use inquiry[9]—even though that position would be destructive of the rights of the news originator.

The lack of clarity in the law concerning how much of the body of a news report can be reproduced or abstracted without creating liability for infringement is exacerbated by doubts as to copyright protection for news headlines. The writing of effective headlines is recognized in the industry as a high art, no less than the writing of ledes. Yet, this is not reflected in the law. "Because few cases address this issue—whether the use of a newspaper's headline and lede, or an important snippet from the story, constitute copyright infringement—[it] remains unsettled." Lindsay Marks, *Can Copyright Save the U.S. News Industry?: Applying the 2016 European Union Proposal to the United States*, 46 AIPLA Q.J. 61, 95 (2018). Those who would argue that headlines are not protectable cite the Copyright Office regulation forbidding the registration of a copyright claim in "words and short phrases," *see, e.g.*, Circular 33, "Works Not protected by Copyright"; 37 CFR § 202.1(a). But this rule does not necessarily govern whether the taking of a headline, considered as part of a news article which is surely protected by copyright, would be

[8] Meltwater did not contest that the extracts it reproduced were sufficiently substantial to infringe absent fair use or one of the several other affirmative defenses it raised. 931 F. Supp. 2d at 559.

[9] This view that TVEyes' Watch function was somewhat transformative because it improves the efficiency with which TVEyes customers can access the original content was sharply contested by the concurring judge. *See* 883 F.3d at 182, 185-86 (Kaplan, D.J., concurring in the result and in parts of the majority opinion) ("Even on the majority's view that TVEyes' Watch function substantially improves the efficiency with which TVEyes customers can access Fox copyrighted broadcasts of possible interest, it does no more than repackage and deliver the original works. It adds no new information, no new aesthetics, and no new insights or understandings. I therefore doubt that it is transformative.").

sufficiently qualitatively (if not quantitatively) substantial to amount to infringement. Some courts outside the United States have answered this question affirmatively.[10]

Even after *TVEyes* and *Meltwater* there remains a perception in the industry, if not a squarely dispositive precedent, that reproduction of headlines and snippets of news shorter than those made available in those cases is either not infringement or fair use. This perception is driven by an outmoded view of the "substantial similarity" test for infringement of fact-based works reflected in *Nihon*, which does not address the advent of short-form journalism: today, many readers have been trained by their constant exposure to limited-character social media posts and the like to satisfy their appetite for news by reading short snippets or headlines rather than long-form articles.

Gradual evolution of case law in ways that may address these concerns will not come in time to save the press from arrogation of news content. To address this problem, the Copyright Office as an integral element of its Publishers' Protection Study should advise on remedial legislation, whether as a matter of copyright law and/or as a right or bundle of rights alongside and complementary to copyright.

B. Limited availability of common law remedies for systematic misappropriation of news threatens to create a law-free zone for free riders

Publications of News Corp subsidiary Dow Jones, such as *The Wall Street Journal* and the *Dow Jones Newswires*, are regular targets of misappropriation of news content. Through that misappropriation, online operators seek to capitalize on the time value of breaking, market-moving business and general news by systematically reproducing and delivering it to their customers at a lower price than Dow Jones is able to charge. It was for situations such as this that the tort of misappropriation was fashioned. Although the world is much changed since the decision in *International News Service v. Associated Press*, 248 U.S. 215 (1918) ("*INS*"), then as now the law was challenged by advances in communications technology—then, telegraphy and telephony, now, the internet—that threatened to deprive news originators of the ability to monetize their output. In *INS*, the Court responded by recognizing as unfair competition the taking of news, whether or not protected by copyright, for sale by a competitor in such a way as to deprive the originator of the timeliness value of the breaking news it had collected.

In the modern digital era, a variety of online operators do so much the same. In the last several years, for example, Dow Jones has filed suit for misappropriation under state law against one company that regularly cut and pasted headlines and short news reports from the *Dow Jones Newswires* into its own website within minutes after they appeared, *Dow Jones & Co., Inc. v. Briefing.com, Inc.*, No. 10 Civ. 3321 (S.D.N.Y. 2010), and another that delivered to traders real time voice renditions ("squawks") of market-moving headlines that appeared in the *Dow Jones*

[10] *Newspaper Licensing Agency Ltd. v. Meltwater Holding BV*, [2010] EWHC 3099 (Ch), [146] (Eng.); *Copiepresse v. Google, Inc*., Tribunal de Première Instance [Civ.] [Tribunal of First Instance], Brussels, Feb. 15, 2007, No. 06/10.928/C (Belg.); Case C-5/08, *Infopaq Int'l A/S v. Danske Dagblades Forening*, 2009 E.C.R. I-06569; *but see Fairfax Media Publ'ns Propriety Ltd. v Reed Int'l Books Australia Propriety Ltd*. (2010) 189 FCR 109, 113 (Aus.) (headlines too short to merit copyright protection).

Newswires over its own audio network, *Dow Jones & Co., Inc. v. Real-Time Analysis & News, Ltd. (d/b/a "Ransquawk")*, 14 Civ. 131 (S.D.N.Y. 2014).

These cases ended successfully for Dow Jones, in the first case by an admission of liability and settlement and in the second by a default judgment, resulting in damages and injunctions against further violations. The efficacy of this tort remedy, however, is diminished because it is a creature of state law. This has at least two complicating consequences. First, it may not be uniform across the 50 states, leading to unnecessarily complex choice-of-law determinations and forum shopping. *See, e.g., Associated Press v. All-Headline News Corp*., 608 F. Supp. 2d 454 (S.D.N.Y. 2009) (applying New York state law where plaintiff was based in New York and defendant in Florida, rejecting claim that Florida law, which assertedly did not recognize misappropriation, should govern). Second, in the modern era the tort must run the gauntlet of federal preemption under section 301 of the 1976 Copyright Act. *See, e.g., Nat'l Basketball Assn. v. Motorola, Inc.*, 105 F.3d 841 (2d Cir. 1997) ("*NBA*").

Unfair competition through the systematic misappropriation of news content deserves to be accepted as adjunct and complementary to copyright in protecting the ability of reporters of news to monetize their output. Such a doctrine is a necessary tool, because often the news content misappropriated is treated by courts as purely factual or the content taken deemed too short to qualify for copyright protection. As the *NBA* decision recognized, the misappropriation tort features five elements: "(i) a plaintiff generates or gathers information at a cost; (ii) the information is time sensitive; (iii) a defendant's use of the information constitutes free-riding on the plaintiff's efforts; (iv) the defendant is in direct competition with a product or service offered by the plaintiffs; and (v) the ability of other parties to free-ride on the efforts of the plaintiff or others would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened." 105 F.3d at 845. This formulation essentially reflects the fact pattern of the *INS* case, and in over 100 years no court has held that, on the facts of *INS*, the plaintiff should not have a remedy. The *INS* decision was animated by advances in technology not foreseen when the then-governing copyright law had been developed. So too, today's technological revolution brought about by the instantaneous digital delivery of information and global interconnectedness cries out for reform of the law. We welcome, as identified in the Notice, the Copyright Office's inquiry into unfair competition law alongside copyright to address the systematic misappropriation of news content.

C. Inability to register dynamic website content on bulk basis

The registration requirement historically served the purposes of providing public notice of copyright claims, identifying associated rightsholders, and centralizing a repository of copyrighted works by building the collection of the Library of Congress in an age when information was recorded on paper. The need for these functions has been steadily—and, in recent years, almost completely—eroded. With the advent of technology allowing instantaneous publication (and republication by infringers), the registration requirement has more predominantly served as an obstacle to copyright protection than as a tool for public information.

The global trend has been to de-formalize the process of obtaining copyright protection. The Berne Convention, to which the United States acceded in 1989, provides that the enjoyment and exercise of intellectual property rights shall not be subject to any formality such as

registration or publication.[11] Nevertheless, the Copyright Act persists in requiring U.S. authors to register their works as a precondition to an infringement suit and to the full suite of statutory protections and remedies, even while it allows an exception for foreign works.[12] U.S. news media organizations, perversely, thus receive less robust copyright protection under U.S. law than their foreign counterparts.

The registration requirement imposes unique challenges in today's world, where a news media organization may generate hundreds of individual works of authorship every day, all of them published in real time. Infringers have access to the same tools that allow this instantaneous publication. Thus, it may be mere minutes after an article has appeared on a website like wsj.com that it is republished verbatim on a blog or other website. Until the copyright claim to such an article is registered, however, the author cannot sue for such infringement. And unless registration was "effective" prior to the infringement or registration is "made" within three months of initial publication, the author is not entitled to statutory damages or attorneys' fees in an enforcement action.[13] This scheme significantly impairs the practical utility of copyright protection. With application processing often consuming several months,[14] news publishers are in jeopardy of being deprived of prompt remedies for infringement even if they apply to register their works immediately after publication. And while the Copyright Office offers an expedited registration option, the "special handling" fee of $800 is greater than the minimum amount of statutory damages the Copyright Act provides.[15] For many news media organizations, whose business models are already subject to unprecedented challenges, it is not economical to apply for expedited registration of their entire corpus of daily content.

Meanwhile, the practical and logistical impediments to comprehensive and timely registration of modern news content are effectively insurmountable as well. There is currently no feasible way to apply for a single group registration of the full contents of a dynamic news media website, on which articles are constantly being published and updated. In many cases, modern journalism is published only on such websites, and does not appear in print at all. One example is Dow Jones's *Mansion Global*,[16] a frequent target of infringers. For a time, a news publisher could register its website as a "database," but several years ago the Copyright Office

---

[11] *See* Berne Convention for the Protection of Literary and Artistic Works, Art. 5.

[12] *See* 17 U.S.C. § 411.

[13] *See* 17 U.S.C. § 412; *Fourth Estate Public Benefit Corp. v. Wall-Street.com,* 139 S. Ct. 881 (2019) (interpreting the Act as requiring that the Copyright Office return a registration certificate, or deny registration, before an infringement suit is filed).

[14] *See Fourth Estate Public Benefit Corp.*, 139 S. Ct. at 892.

[15] *See* 17 U.S.C. § 504.

[16] https://www.mansionglobal.com/.

announced that this practice would no longer be permitted.[17] Mobile apps create a similar concern. Thus, if an infringer has cherry-picked and reproduced several hundred news articles from a news website, the publisher would have to register each article separately—and probably pay a special handling fee for each registration—to bring a timely suit for infringement of those works.[18] Indeed, a recent Dow Jones complaint for copyright infringement explained that while Dow Jones registers its print content as a matter of course, it is "unable to register its online content in similar fashion because there is no comparable method for registration made available by the Copyright Office for online content."[19] The result has been to multiply the costs of policing copyright infringements, to leave organizations with no feasible means to access all of the statutory remedies for infringement of their online news content, and to leave infringers free to pick and choose from works among that content without fear of significant or immediate consequence.

The availability of statutory damages and an award of attorneys' fees is vital to the meaningful enforcement of copyright in news content. Without those potential remedies, a plaintiff is left to seek the infringer's profits or its own damages. And while the harm caused by infringement writ large is enormous, infringers are emboldened by the fact that direct financial injury resulting from any individual act of copying a single article is often difficult to prove. While there is an established market for reprints of individual articles, the vast majority of news organizations' revenue comes not from the sale of individual works but from subscription fees or from advertisers courting the attention of an exclusive audience to which only the publisher can provide access. Infringement grievously undermines these streams of revenue, but not in a way that is readily quantifiable or redressable in a judicial proceeding about individual article-length works.

In sum, a generation after the popularization of the internet, U.S. copyright registration law and practice makes it beyond difficult for news publishers to acquire meaningful copyright protection for their output. To make copyright protection more than just an empty promise to 21st-century American news media organizations, Congress and the Copyright Office must make it practical and economical to access the remedies for infringement, and in particular for infringement of content on news websites that are updated by the minute. The U.S. copyright

---

[17] *See* U.S. Copyright Office, Compendium of U.S. Copyright Office Practices § 1002.6 ("Websites may contain databases, but they are not considered databases for the purpose of copyright registration.").

[18] In Circular 66, the Copyright Office explains that "[a]s a general rule, you should submit a separate application for each component work appearing on the website, although it is possible to register multiple works on one application if they qualify for one of the Office's special registration accommodations." U.S. Copyright Office, Circular 66: Copyright Registration of Websites and Website Content, available at https://www.copyright.gov/circs/circ66.pdf. The "special registration accommodations" made available by the Copyright Office are "limited" in nature, and in many cases apply only to works that have not yet been published—a requirement that cannot practically be satisfied by news organizations that maintain dynamic websites that are updated multiple times throughout the day. *See* U.S. Copyright Office, Circular 34: Multiple Works, available at https://www.copyright.gov/circs/circ34.pdf.

[19] Compl. ¶ 45, *Dow Jones & Company, Inc. v. Juwai Ltd.*, 21-cv-7284 (S.D.N.Y. Aug. 30, 2021).

registration regime should therefore be reconsidered and either abrogated or reformed so as not to be an obstacle to enforcement of copyright in news content.

D. Efficacy of the remedy for circumvention of access controls under Section 1201

Among other things, the Digital Millennium Copyright Act of 1999 made it an actionable violation to "circumvent a technological measure that effectively controls access to a work protected under [the Copyright Act]."[20] This and other protections offered by the DMCA are a critical modern supplement to a statutory framework designed for a different era.

Nevertheless, the DMCA has not always been applied consistently. For example, the plain text of the anti-circumvention provision suggests that unauthorized use of another entity's login information to evade a paywall is prohibited. Nevertheless, some courts have held that the statute does not reach the use of "a password intentionally issued by plaintiff to another entity."[21] Other courts have taken the opposite view.[22] Permitting such evasion makes it difficult for an online publisher to detect and police redistribution.

To resolve this split in authority, Congress should clarify that circumvention of an organization's paywall through the use of another's password violates the DMCA. Dow Jones has observed that in recent years, infringers are building business models around the systematic appropriation and widespread dissemination of copyrighted content by using a small number of authorized accounts to gain access to online news content. This is precisely the kind of practice that the DMCA was intended to prevent. While the misuse of login credentials and resulting redistribution likely is also a breach of website terms of service, stronger remedies are needed.

Of equal importance, some courts, led by the Federal Circuit, have engrafted onto section 1201(a) a requirement that "[a] copyright owner . . . prove that the circumvention of the technological measure either 'infringes or facilitates infringing a right protected by the Copyright Act.'"[23] Other courts have rejected the Federal Circuit approach and, in accordance with the text of the DMCA, held that an actionable violation of section 1201(a) does not require a "nexus to infringement."[24] Decisions in the former category depart from the statutory text and deprive the DMCA anti-circumvention rule of much of its force. The Copyright Office in its Publishers'

---

[20] 17 U.S.C. § 1201.

[21] *See I.M.S. Inquiry Mgmt. Sys., Ltd. v. Berkshire Info. Sys., Inc.*, 307 F. Supp. 2d 521, 532–33 (S.D.N.Y. 2004).

[22] *See, e.g., Actuate Corp. v. Int'l Bus. Machines Corp.*, 2010 WL 1340519, at *9 (N.D. Cal. Apr. 5, 2010).

[23] *Storage Tech. Corp. v. Custom Hardware Eng'g & Consulting, Inc*., 421 F.3d 1307, 1318-19 (Fed. Cir. 2005), quoting *Chamberlain Group, Inc. v. Skylink Technologies, Inc.,* 381 F.3d 1178, 1203 (Fed. Cir. 2004).

[24] *MDY Indus., LLC v. Blizzard Ent., Inc.*, 629 F.3d 928, 949-52 (9th Cir. 2010), *as amended on denial of reh'g* (Feb. 17, 2011).

Protects Study should address this split in authority and the Federal Circuit approach should be rejected to align with the objectives and intent of the DMCA's anti-circumvention rule.

E. Congress' failure to provide for a constitutionally acceptable cause of action against state infringers has created another gap in protection

In *Allen v. Cooper*, 140 S. Ct. 994 (2020), the Supreme Court held unconstitutional the Copyright Remedies Clarification Act, in which Congress had amended 17 U.S.C. § 511(a) to declare state governments and their instrumentalities liable for copyright infringement.

There is no sound policy reason to exempt state agencies from copyright liability. Making clear that state actors must respect the copyrights of publishers will go a long way to protect the legal rights and economic incentives of the press, as a recent incident involving Dow Jones and other news publishers illustrates. In the years immediately prior to *Allen*, Dow Jones was a principal victim of what must stand as the most egregious exemplar of copyright infringement by a state. In June 2017, a blogger active in the financial news arena posted an article[25] revealing that the California Public Employees' Retirement System ("CalPERS"), a California state agency that administers the nation's largest pension fund, was maintaining a publicly accessible website that featured daily postings of articles from, among others, Dow Jones's premier publications, including *The Wall Street Journal*, *Barron's*, *MarketWatch*, and *Dow Jones Newswires*. Dow Jones had never authorized this activity. The sheer number of Dow Jones articles copied between 2009 and 2017 was staggering: approximately 9,000 full-text articles from *The Wall Street Journal*, 257 from *Barron's*, and over 560 items from other Dow Jones publications. Among the pirated pieces were numerous articles taken from *Dow Jones Newswires*, a high-value suite of news services for financial and other business professionals featuring breaking, exclusive, and often market-moving news.

CalPERS disseminated more of Dow Jones's copyrighted reports than it did those of any other publisher, but Dow Jones was not the only victim. Among the republished full-text articles found on the CalPERS website were approximately 6,700 articles taken from *The New York Times*, 5,400 from *The Los Angeles Times*, over 3,100 from *The Sacramento Bee*, and over 1,500 from *The Washington Post*. All told, CalPERS had reproduced some 53,000 separate articles from approximately 4,500 publishers over an eight-year span—including articles from essentially every major daily newspaper, business periodical, and cable news network. Dow Jones later learned that CalPERS had also compiled these selected articles into a daily email that it sent to approximately 200 senior executives and other recipients, both within and outside of CalPERS, who could then forward the content at will to anyone they chose. By doing so, CalPERS created a curated daily newsfeed to serve the needs of those on its distribution list for business and financial news of potential importance to CalPERS and pensions generally—a natural audience for Dow Jones's publications. CalPERS thus competed directly with Dow

[25] Yves Smith, "CalPERS Internal News Site Ignores Unfavorable Stories, Steals Copyrighted Material," Naked Capitalism (June 9, 2017), available at https://www.nakedcapitalism.com/2017/06/calpers-internal-news-site-ignores-unfavorable-stories-steals-copyrighted-material.html.

Jones (and with the many other news originators whose output it misappropriated) to serve that demand, and diverted substantial paying business from the authorized publications.

CalPERS resisted Dow Jones' copyright claim by invoking sovereign immunity. Eventually, Dow Jones and several other publishers entered into settlement agreements with CalPERS, but for sums well below what would have been warranted in the absence of a threatened sovereign immunity defense. Dow Jones has subsequently detected at least one other infringer of *Wall Street Journal* articles who has claimed sovereign immunity as a defense after *Allen.*

The *Allen* decision left open the possibility that Congress could validly abrogate state immunity for copyright infringement. The Copyright Office should support this objective, and Congress should do so.

F. Ingestion for machine learning and data mining

The science of artificial intelligence ("AI") has advanced at an accelerating pace in recent years. Tech entities have long sought to use online content for natural language processing, machine learning, and data mining purposes. Those businesses have benefited from the vast amount of text that is "available en masse thanks to the Internet."[26] Among such content is an almost unlimited supply of news articles and text generated from publishers.[27] Indeed, news content, particularly from trusted sources, has proven to be a highly attractive and useful input for purposes of machine learning and the development of AI-based tools.[28]

For some time, media entities have identified this market as one in which their proprietary content has particular value and have curated and made available annotated corpora of their published news reporting for the specific purpose of training AI. This has been done by,

[26] Colin Raffel, et al., *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*, J. of Machine Learning Res., June 2020, at 2, available at https://arxiv.org/pdf/1910.10683.pdf ("[U]nsupervised pre-training for NLP is particularly attractive because unlabeled text data is available en masse thanks to the Internet—for example, the Common Crawl project produces about 20TB of text data extracted from web pages each month.").

[27] *Id.* at 25 ("Recent work has used text data extracted from news websites"); Jingqing Zhang, et al., *PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization*, Proceedings of the 37th International Conference on Machine Learning, 2020, at 4, available at https://arxiv.org/pdf/1912.08777.pdf ("NEWSROOM … is a large dataset containing 1.3M article-summary pairs written by authors and editors in the newsrooms of 38 major publications between 1998 and 2017.").

[28] *See id.*; Rubik's Code, *Top 23 Best Public Datasets for Practicing Machine Learning* (July 19, 2021), available at https://rubikscode.net/2021/07/19/top-23-best-public-datasets-for-practicing-machine-learning/ (listing BBC News Datasets among top training sets).

for example, the copyright holders of *The Wall Street Journal*,[29] *The New York Times*,[30] and the Reuters News Service.[31] The Linguistic Data Consortium catalogue lists hundreds of such corpora available for license.[32]

Taking a step further, some companies have started to use artificial intelligence to assist in the process of generating news articles and news content.[33] The prospect looms of tech companies using news content generated by the efforts of traditional journalistic enterprises to train their computers to write news stories that extract the non-copyrightable facts from human-created news reports and then compete directly with the human-made publications.

The legal status of such commercial uses warrants clarification and guidance. While the ingestion of vast amounts of news reporting results in a *prima facie* infringement of the reproduction right, AI operators are likely to claim fair use because they input, but in many cases do not output, a work that is substantially similar to the original. Some of the uses to which tech entities put machine learning involve scientific research and for that reason they will also claim fair use, although the strength of that claim is questionable. *See American Geophysical Union v. Texaco Inc.*, 60 F.3d 913 (2d Cir. 1995) (extensive internal photocopying of scientific journal articles for use in research not a fair use). Other end uses of machine learning are purely commercial, although they may or may not directly compete with the copyright-protected texts

---

[29] Linguistic Data Consortium – BLLIP 1987-89 WSJ Corpus Release 1, available at https://catalog.ldc.upenn.edu/LDC2000T43 (last visited Jan. 6, 2020). Dow Jones also offers a curated corpus of news reporting tagged for machine learning purposes, known as DNA. *See* https://www.dowjones.com/news-you-can-use/ ("DNA provides API-based access to 8,000 (and growing) content sources in the form of 'snapshots' of rich archives and real-time 'streams' where data is delivered continuously. This content is licensed for text-mining and machine-learning use cases").

[30] Linguistic Data Consortium – The New York Times Annotated Corpus, available at https://catalog.ldc.upenn.edu/LDC2008T19 (last visited Jan. 6, 2020). For a description of the corpus, see this blog post announcing and explaining the corpus: Jacob Harris, *Fatten Up Your Corpus*, NYT Open (Jan. 12, 2009), available at https://open.blogs.nytimes.com/2009/01/12/fatten-up-your-corpus/ ("Available for noncommercial research license from The Linguistic Data Consortium (LDC), the corpus spans 20 years of newspapers between 1987 and 2007 (that's 7,475 issues, to be exact). This collection includes the text of 1.8 million articles written at The Times . . . .").

[31] *See, e.g.*, David D. Lewis, Reuters-21578 Text Categorization Test Collection Distribution 1.0 README file (Sep. 26, 1997), available at https://perma.cc/V7JJ-CNVW. This corpus consists of the contents of the Reuters newswire for 1987.

[32] Linguistic Data Consortium, LDC Catalog, https://catalog.ldc.upenn.edu/ (last visited Jan. 6, 2020).

[33] Mack DeGeurin, *A Startup Media Site Says AI Can Take Bias Out of News*, Vice (April 4, 2018), available at https://www.vice.com/en/article/zmgza5/knowhere-ai-news-site-profile ("AI has also started writing rudimentary news articles and assisting reporters, but a new startup launched Wednesday says it will use AI to publish breaking news about a wide variety of topics."); Farhad Manjoo, Opinion, *How Do You Know a Human Wrote This?*, N.Y. Times (July 29, 2020), available at https://www.nytimes.com/2020/07/29/opinion/gpt-3-ai-automation.html (describing GPT-3, a "powerful" language processing model "trained on an enormous corpus of text").

used to generate those products and services. Finally, tech entities will claim that they are doing no more than extracting uncopyrightable facts from the ingested texts, but this too is questionable: sophisticated modern machine learning techniques focus on learning from the expressive content of the ingested works. As one commentator has observed, "[m]achine learning gives computers the ability to derive valuable information from the way authors express ideas. Instead of merely deriving facts about a work, they may be able to glean value from a work's expressive aspects; as a result, these uses of machine learning may no longer qualify as non-expressive in character." Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 Colum. J. L. & Arts 45, 507 (2017).

Publishers of original content at great cost whose works are essential to the employment of AI processes for commercial purposes should be compensated for such uses. The EC 2019 Copyright Directive addresses this issue head-on, protecting from liability the ingestion of content for machine learning and data mining for purely scientific research purposes, thus leaving commercial uses subject to liability. In this respect European law has begun to address cutting edge issues at the intersection of copyright law and modern technology in ways that U.S. law has not yet done.

As this comment has discussed, technological developments of recent decades have far outpaced the evolution of copyright protection. But despite the disruptive implications of AI for news reporting and intellectual property generally, those developments may soon prove to have been a mere prelude. The time to adopt policies that create and protect incentives to generate original news content in the face of previously unthinkable technological capacity is now, before the field has been irrevocably altered by automation.

G. The unlevel playing field

In the previous sections we have outlined gaps in intellectual property protection that have tilted the balance too far in favor of uses by certain actors of original news content at the expense of originators. But any recalibration of copyright and related law to address those gaps likely will prove ineffectual if the current imbalance in negotiating power between publishers and republishers of news content is not remedied. We have discussed how the technological changes that have enabled swift and virtually cost-free online republication and the concomitant failure of copyright law to keep up with those changes have devastated the news industry. But even conferring new rights and updating those that now exist will not remedy this situation if publishers do not have the practical ability to require republishers to negotiate licenses on fair and reasonable terms, and to prevent those republishers from playing one publisher off against another

While this issue presents questions of competition law and policy that may go beyond the institutional remit of the Copyright Office, it needs to be recognized as an integral element of the concerns animating the Publishers' Protection Study. There is now pending in Congress a bill to immunize news publishers from liability for a period of time to collectively negotiate with republishers over the use of news content.[34] The Copyright Office should support that bill in its

[34] The Journalism Competition and Preservation Act of 2021, S. 673 (117th Cong.).

Publishers' Protection Study. In addition, the Copyright Office should support activities in Congress to review the current need for, and potential reforms to, section 230 of the Communications Decency Act, which generally immunizes website operators from liability arising from content posted by third parties. While many have applauded section 230 as necessary to cultivate the growth of the nascent internet in the 1990s, it has fostered the proliferation of unedited, unsupervised misinformation online. Limiting the protection of section 230 may provide incentives for republishers of third party content to contract with originators of news content, who professionally produce and stand behind their reporting.

4. Conclusion

News Corp appreciates the attention that the Copyright Office is devoting to the profound and possibly fatal threats faced by the news media and ways to update copyright and other relevant bodies of law to address those challenges. To restore effective legal protection for news publications, four things are needed: (1) fortification of existing copyright protections for the evolving digital age, (2) recognition of additional rights necessary to monetize original news content, (3) procedural reforms, both in the Copyright Office and in the courts, to enable frictionless enforcement of rights in news content, and (4) a level playing field to enable publishers to meaningfully negotiate licensing of those rights with redistributors. In this comment News Corp has sought to identify key areas where one or more of those necessities is currently lacking.

It is imperative that U.S. law and practice be updated without delay to forestall further decline of the news media. Inevitably, the solutions will include not only modifications to the Copyright Act but also reliance on adjacent and other bodies of law. For example, within Title 17, Congress has enacted copyright-like protection for semiconductor mask works and for vessel hulls and other original designs of useful articles (17 U.S.C., ch. 9 and 13), and the European Union has promulgated ancillary rights to protect news publications from misappropriation. EC 2019 Copyright Directive, Art. 15. If we want to protect a robust, vibrant news media industry that produces accurate, quality content, law and policy-makers in the United States must evaluate a wide range of options and take decisive action.

We look forward to working with the Copyright Office toward solutions.