

ALL THE NEWS THAT’S NOT FIT TO SCRAPE:

The Associated Press Granted Summary Judgment on Its Copyright Claims Against Internet News Aggregator Meltwater

James G. Snell and Derek Care

In 1918, the Associated Press (“AP”) brought suit against the International News Service (“INS”) in connection with its practice of redistributing the AP’s breaking news reports using the most advanced communications technology of the time – the telegraph. That suit led to a seminal decision by the United States Supreme Court, *International News Service v. Associated Press*, holding that the INS could be temporarily restrained from redistributing the facts contained in AP’s breaking news reports (*i.e.*, “hot news”), based in part on the Court’s concern that the AP have an opportunity to obtain a just profit from its news reporting activities and thus have an incentive to continue with such activities, which the Court found benefit the public.¹

Nearly a century later, the AP has obtained a legal victory against a company who made use of the AP’s content using significantly more advanced technology. On March 21, 2013, in connection with a case captioned *Associated Press v. Meltwater Holdings U.S., Inc.*, Judge Denise Cote of the United States District Court for the Southern District of New York granted summary judgment on the AP’s claims for copyright violations by defendants (collectively “Meltwater”) who “scraped” news articles available on the websites of AP licensees, and distributed excerpts of and links to those articles to paid subscribers to Meltwater’s online news monitoring service.² This decision – which, like *International News Service*, cited the need to preserve incentives for parties to undertake the effort and expense of gathering and disseminating news, to the significant benefit of the public – presents one of the most comprehensive judicial discussions to date of the legal issues raised by Internet news aggregation. It also raises a number of issues with potential implications for both online content owners and entities whose business involves the use of web-crawling and scraping tools, including for purposes of content aggregation and data analytics. The Meltwater decision and its potential implications are discussed below.

MELTWATER AND THE AP’S COPYRIGHT CLAIMS

As discussed in the court’s decision, Meltwater is a Norwegian company that, since 2005, has offered subscription-based news monitoring services to business, non-profits and government

¹ See generally 248 U.S. 215 (1918). The Supreme Court’s recognition of a claim for “hot news” misappropriation was based on federal common law, which the Supreme Court later held does not exist. See generally *Erie R. Co. v. Tompkins*, 304 U.S. 64 (1938). Today, “hot news” misappropriation is recognized as a cause of action only under the laws of five states, and has been held to be preempted by the Copyright Act only in circumstances essentially identical to that which were before the Supreme Court in *International News Service*. See *Barclays Capital Inc. v. Theflyonthewall.com, Inc.*, 650 F.3d 876, 902-04 (2d Cir. 2011) (discussing factual circumstances that might give rise to a “[*International News Service*]-type claim [that] might survive preemption”).

² --- F. Supp. 2d ---, No. 12 Civ. 1087 (DLC), 2013 WL 1153979 (S.D.N.Y. Mar. 21, 2013).

agencies in the United States. Meltwater’s service is based on its use of automated software that “crawls” (*i.e.*, frequently visits) more than 162,000 online news sources and “scrapes” (*i.e.*, indexes or copies to a database) news content from those websites. The scraped websites include those of the AP’s licensees. From its database of scraped content, Meltwater generates customizable reports that it sends to its subscribers containing excerpts of articles containing keywords selected by subscribers as well as links to those articles. While the amount of copying varies – ranging, for example, from 4% to 61% of AP articles scraped from its licensees’ websites – Meltwater’s reports generally include the headline and the first 300 characters of the articles, as well as a 150-character excerpt of text surrounding the relevant keyword. As a result, Meltwater’s reports generally reproduce, at a minimum, the “lede” of the article, *i.e.*, an introductory sentence or section conveying the key facts or message of the article. Meltwater also provides its subscribers with analytical tools concerning the news it monitors, such as charts and graphs indicating which words, topics and geographical locations are “trending” in media coverage. Meltwater also allows its subscribers to conduct searches of its news database.³

In 2012, the AP filed suit against Meltwater in the United States District Court for the Southern District of New York, alleging copyright infringement, “hot news” misappropriation, and related claims in connection with Meltwater’s alleged copying and distribution of AP articles scraped from the websites of the AP’s licensees. The AP sought damages and injunctive relief in connection with its claims. After targeted discovery relating to the AP’s copyright claims, the AP moved for summary judgment on those claims and as to Meltwater’s affirmative defenses to those claims based on the “fair use” doctrine, implied license and equitable estoppel.⁴ In its March 21 decision, the court granted each of AP’s motions, finding that the undisputed evidence demonstrates that Meltwater violated the AP’s copyrights and has no valid defense to the AP’s claims. The court’s decisions on these motions are discussed below.

MELTWATER’S “FAIR USE” DEFENSE

Meltwater did not contest either the AP’s evidence that the AP possesses valid copyrights in the articles Meltwater scraped from the AP licensees’ websites, or that Meltwater copied protected elements of those articles. Meltwater instead argued that its copying is a fair use of the articles because Meltwater’s news monitoring service functions as a search engine or “information-

³ *Id.* at *3-6.

⁴ The AP also moved for summary judgment as to Meltwater laches and copyright misuse defenses, which motions the court granted in its March 21 decision, except that the court denied without prejudice the AP’s motion for summary judgment as to Meltwater’s laches defense as concerned the AP’s claim for injunctive relief so that discovery necessary to that motion could be conducted. *Id.* at *26-28. The parties also cross-moved for judgment on the pleadings as to the AP’s other claims, including AP’s “hot news” misappropriation claim, and as to counterclaims by Meltwater for libel, tortious interference, and declaratory judgment of safe harbor under the Digital Millennium Copyright Act, 17 U.S.C. § 512(c). *Id.* at *8 n.7. The court did not address those motions in its March 21 decision, suggesting that such claims would be addressed in a forthcoming decision. *Id.*

location tool” by providing its subscribers with limited amounts of the copied content in response to user searches, along with links to the source of that content.⁵

The court rejected this argument, holding that a balancing of the four non-exclusive “fair use” factors set forth in the Copyright Act – *i.e.*, (1) the purpose and character of the use of the copyright work, (2) the nature of the work, (3) the amount and substantiality of the use of the work, and (4) the effect of the use upon the potential market for the work – weigh decisively against a finding of fair use.⁶ In so holding, the court focused on the first, third and fourth factors, specifically as follows.⁷

THE COURT HELD THAT MELTWATER’S USE OF THE AP’S ARTICLES IS NOT TRANSFORMATIVE

The court devoted the bulk of its fair use analysis to consideration of Meltwater’s argument that its copying of the AP’s articles is transformative, including because Meltwater purportedly used those articles in connection with providing search engine services, and because it used the information it scraped to develop analytics relating to the online news sources it monitors. The court rejected each of these arguments, finding that the purpose and character of Meltwater’s use of the AP’s articles weigh strongly against a fair use finding.

First, the court found that Meltwater’s news monitoring service does not transform the AP’s content in any way. Instead, the court held that Meltwater functions as a high-tech clipping service that “uses its computer programs to automatically capture and republish designated segments of text from news articles, without adding any commentary or insight in its News Reports.” Such conduct, the court held, constitutes “undiluted use” of the AP’s protected content.⁸

Moreover, while “new and reporting” are specifically identified in the preamble of the Copyright Act as grounds for the fair use doctrine, the court found that this fact did not render Meltwater’s use of the AP’s article transformative since “[t]he news reporting and research upon which Meltwater relies was not done by Meltwater but by AP; the copyrighted material that Meltwater has taken is the news reporting and research that AP labored to create.”⁹ Such free-riding, the court found, harms not only the AP but also the general public, which benefits from gathering

⁵ *Meltwater*, 2013 WL 1153979, at *9.

⁶ *Id.* at *11, citing 17 U.S.C. § 107.

⁷ With respect to the second fair use factor, *i.e.*, the nature of the copyrighted work, the court briefly noted that the AP’s articles (1) are factual, and therefore more subject to fair use than fictional works, and (2) were already published by the time of Meltwater’s copying. This factor was therefore “at most neutral on the question of fair use, and should be weighed in favor of finding fair use.” *Meltwater*, 2013 WL 1153979, at *17 (citation omitted).

⁸ *Id.* at *12.

⁹ *Id.*

and dissemination of news by entities such as the AP and would be harmed if such entities were not able to obtain just compensation for their efforts.¹⁰

Second, the court rejected Meltwater’s argument that its use of the AP’s articles is transformative because Meltwater functions as a search engine, which courts have recognized as making transformative use of copyrighted works displayed in search results since such results facilitate the public’s access to information on the Internet.¹¹ As a factual matter, the court found that Meltwater *does not* function like traditional search engines (which are available to the general public, and conduct searches of information throughout the Internet), but instead acts as a high-tech clipping service that gathers information only from certain online news sources and provides excerpts of that information to a small number of paid subscribers.¹² Moreover, whereas the use in search engine results of small, low-resolution copies of high-quality photographs has been found to be a fair use insofar as the former is not a substitute for the latter, the court found that Meltwater reproduces the key elements of the AP’s articles, including the article headline and lede, and thus provides a substitute for the AP’s article.¹³ The court cited as evidence confirming this finding the fact that Meltwater’s subscribers followed links to the excerpted articles less than 1% of the time, whereas other news aggregation services have estimated click-through rates as high as 56%.¹⁴ Accordingly, the fact that Meltwater “use[s] an algorithm to crawl over and scrape content from the Internet is surely not enough to qualify as a search engine engaged in transformative work.”¹⁵

Further, the court found that Meltwater’s “search engine” argument was ultimately beside the point, since “using the mechanics of search engines to scrape material from the Internet and provide it to consumers in response to their search requests does not immunize a defendant from the standards of conduct imposed by law through the Copyright Act, including the statutory embodiment of the fair use defense.”¹⁶ Thus, even if Meltwater did function as a

¹⁰ *Id.* at *13 (“Investigating and writing about newsworthy events is an expensive undertaking and enforcement of the copyright laws permits AP to earn the revenue that underwrites this work. Permitting Meltwater to take the fruit of AP’s labor for its own profit, without compensating AP, injures AP’s ability to perform this essential function of democracy.”).

¹¹ *See Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003).

¹² *See Meltwater*, 2013 WL 1153979, at *15 (distinguishing *Perfect 10* and *Kelly*). In *Perfect 10* and *Kelly*, the Ninth Circuit held that an Internet search engine’s display of low-resolution “thumbnail” copies of high-resolution copyrighted photos were a transformative, fair use of the original works because the search engines facilitated the public’s access to information on the Internet, and because the “thumbnails” at issue did not serve the same purposes as the original photos. *Perfect 10*, 508 F.3d at 1154-55, 1165, 1168; *Kelly*, 336 F.3d at 815-20.

¹³ *Id.* at *14.

¹⁴ *Id.* at 14 & n.13.

¹⁵ *Id.* at *15.

¹⁶ *Id.*

search engine, it nonetheless had to demonstrate, based on a balancing of the four factors, that its excerpting of the AP's articles is a fair use.

Third, the court rejected Meltwater's argument that its copying of the AP's articles is transformative because it also uses the content that it scrapes from online news sources to generate analytics relating to news trends and coverage. The court found that those services, which the AP did not challenge, are separate from Meltwater's excerpting of the AP's content, and thus could not render that excerpting transformative.¹⁷

THE COURT HELD THAT MELT WATER COPIED QUANTITATIVELY AND QUALITATIVELY SIGNIFICANT PORTIONS OF THE AP'S ARTICLES

The court held that a quantitative and qualitative assessment of Meltwater's copying weighed against a finding of fair use. With respect to the quantitative aspect, the court relied on evidence showing that Meltwater copied word for word as much as 61% of the AP's articles – an amount far in excess of the minimum amount of copying that the Court of Appeals for the Second Circuit has recognized as weighing against a finding of fair use, i.e., 8% copying.¹⁸ With respect to the qualitative aspect, Meltwater copied the crucial elements of the AP's articles, including the lede – the most important sentence of an AP article, and that which requires the most journalistic skill to produce.¹⁹ The court also emphasized Meltwater's failure to put forth evidence showing that its nearly 500-character excerpts contained no more of the AP's content than was necessary to perform a search engine service, the purpose by which Meltwater sought to justify its use of the AP's content. In contrast, the AP put forth evidence showing that search engines generally display far briefer excerpts of news articles in search results than does Meltwater, e.g., news alerts put out by other news aggregation services averaging 207 characters in length.²⁰

THE EFFECT OF THE USE ON THE POTENTIAL MARKET OR VALUE FOR THE AP'S ARTICLES

The court found this factor weighed strongly against a finding of fair use in light of the court's findings that (1) the AP and Meltwater compete in the same commercial space for the same customers; (2) Meltwater obtains an unfair commercial advantage by profiting off the AP's news-gathering efforts without paying any license fees to the AP; and (3) Meltwater cheapens the value of the AP's services by competing with companies that do obtain licenses from the AP. Accordingly, the court held that Meltwater obtains an "unfair commercial advantage in the

¹⁷ *Id.* ("The fact that Meltwater also offers a number of analysis tools does not render its copying and redistribution of article excerpts transformative.")

¹⁸ 2013 WL 1153979, at *18 ("The Second Circuit has found that copying as little as eight percent of the original work may tip this factor against a finding of fair use."), citing *Iowa State Univ. Research Found., Inc. v. Am. Broad. Co.*, 621 F.2d 57, 61-62 (2d Cir. 1980) (finding that 8% copying of protected work weighed against fair use).

¹⁹ *Id.*

²⁰ *Id.* at *14 n.16, *18.

marketplace and directly harm[s] the creator of expressive content protected by the Copyright Act.”²¹

A BALANCING OF THE RELEVANT FACTORS, ACCORDING TO THE COURT, WEIGHED DECISIVELY AGAINST FAIR USE

Considering these factors in the aggregate, the court held that Meltwater failed to raise a material question of fact as to the viability of its fair use defense. Rather, the undisputed evidence demonstrated that Meltwater functions as a non-transformative high-tech clipping service.²² The court further held that rejecting Meltwater’s fair use defense “will further the ultimate aim of the Copyright Act, which is to stimulate the creation of useful works for the public good.”²³ Finally, the court re-emphasized its rejection of Meltwater’s argument that its services were immunized from copyright infringement insofar as they are analogous to those of search engines. Not only did the court find that Meltwater’s news monitoring service is functionally dissimilar to that of search engines – specifically because that service does not enhance the public’s access to information generally, but instead helps only paying customers obtain content from certain online news sources – but even if the service was similar to a search engine, Meltwater failed to demonstrate that its specific copying of the AP’s content for purposes of those services was a fair use.²⁴

THE AP DID NOT GRANT MELTWATER AN IMPLIED LICENSE TO USE ITS CONTENT

The court granted the AP’s motion for summary judgment on Meltwater’s claim that the AP’s conduct created an implied license for Meltwater to copy its articles, finding that there was no “meeting of the minds” or pre-existing relationship between the AP and Meltwater capable of suggesting the existence of an implied license.²⁵

Meltwater based its implied license defense on the fact that the AP does not require its licensees to employ the robots.txt protocol – *i.e.*, programming language that a website can deploy to instruct cooperating web crawlers not to access all or part of the website – to deter unauthorized scraping of the AP’s articles.²⁶ The court emphatically rejected this premise, stating that “[w]hat Meltwater is suggesting would shift the burden to the copyright holder to prevent unauthorized use instead of placing the burden on the infringing party to show it had properly taken and used content.”²⁷ The court further held that the AP’s failure to require its licensees to use the robots.txt protocol did not imply a meeting of the minds between the AP and Meltwater – and indeed, a contrary finding would suggest an implied license to any web

²¹ *Id.* at *21.

²² *Id.*

²³ *Id.*

²⁴ *Id.*

²⁵ *Id.* at *23.

²⁶ *Id.*

²⁷ *Id.* at *24.

crawler regardless whether welcome or unwanted.²⁸ The court also deemed Meltwater’s argument “impractical” because (1) the AP’s licensing agreement allows for scraping by various authorized online news sources; and (2) Meltwater reserved the right to ignore the robots.txt protocol.²⁹

Finally, the court found that the public policy of encouraging “openness” of information on the Internet requires rejection of Meltwater’s argument that failure to use the robots.txt protocol granted an implied license:

The protocol is a helpful innovation that gives instructions to co-operating crawlers. But, in the interest of openness on the Internet, one would expect it to be used only when it is in the clear interest of the website to broadly limit access. It is fair to assume that most Internet users (and many owners of websites) would like crawlers employed by search engines to visit as many websites as possible, to include those websites in their search results, and thereby to direct viewers to a vast array of sites. Adopting Meltwater’s position would require websites concerned about improper copying to signal crawlers that they are not welcome.³⁰

Accordingly, the court held that Meltwater’s implied license defense failed as a matter of law.³¹

THE AP IS NOT EQUITABLY ESTOPPED FROM ASSERTING COPYRIGHT INFRINGEMENT AGAINST MELTWATER

Meltwater argued that the AP is equitably estopped from asserting copyright claims against Meltwater because (1) the AP did not require its licensees to use the robots.txt protocol or protect AP content behind pay walls; and (2) the AP did not contact Meltwater prior to filing its lawsuit to demand that it cease-and-desist its use of the AP’s content. The court rejected this defense as a matter of law – finding that the AP had no duty to undertake any of these actions,

²⁸ *Id.*

²⁹ *Id.*

³⁰ *Id.*

³¹ In reaching this decision, the court distinguished the circumstances before it from cases in which courts held that copyright claims could not be stated against search engines who had created “cache” copies of the plaintiff’s website where the plaintiff knew that the search engines would create such cache copies if the protocol were not deployed, but nonetheless refrained from deploying that protocol. *Id.* at *25 (discussing, *inter alia Parker v. Yahoo!, Inc.*, No. 07 Civ. 2757, 2008 WL 4410095 (E.D. Pa. Sept. 25, 2008) (finding that plaintiff granted an implied license to search engines to create cache copies of his website where he knew that defendants would create cache copies of his website in the absence of the robots.txt protocol, which plaintiff did not deploy)). The court specifically distinguished *Parker* on grounds including that Meltwater reserved the right to ignore the robots.txt protocol even if deployed, whereas the search engines in *Parker* had a policy and practice of complying with the protocol if deployed. *Meltwater*, 2013 WL 1153979, at *25.

and that Meltwater could not have interpreted these actions as indicating that its use of the copyrighted content was authorized.³²

Further, the court noted that the undisputed evidence strongly suggests that Meltwater's conduct was not authorized. *First*, most if not all of the AP's licensees' websites are subject to terms-of-use that prohibit "commercial use of the website's content."³³ *Second*, the AP brought suit in 2007 against an online news aggregator that allegedly engaged in conduct similar to that of Meltwater.³⁴ *Third*, the AP had, prior to filing its lawsuit, publicly announced an initiative to protect online news from misappropriation. Reasonable diligence would have put Meltwater on notice of this evidence and would have dispelled any notion that its copying of the AP's content was authorized.³⁵

IMPLICATIONS FOR CONTENT OWNERS, NEWS AGGREGATORS, AND WEB CRAWLERS AND SCRAPERS

As is typically true of copyright cases involving claims of fair use, the holding in *Meltwater* that defendants' excerpting of the AP's content is not a fair use is highly specific to the factual circumstances before the court. Accordingly, *Meltwater* should not be interpreted as establishing a bright-line rule that a content aggregator does not engage in a fair use of copyrighted material where it distributes excerpts of and links to those materials. This is particularly the case because the *Meltwater* court placed great emphasis on the ways that Meltwater's news monitoring services differ from those of other news aggregation services, including that (1) Meltwater's nearly 500-character excerpts are approximately twice as long as that offered by other news aggregators; (2) Meltwater's users follow links to the excerpted articles less than 1% of the time, whereas other news aggregation services have a click-through rate of approximately 50%; and (3) other news aggregators license content from the AP, putting them at a competitive disadvantage with respect to Meltwater (and thereby undermining the value of the AP's license to the other news aggregators) insofar as Meltwater uses the AP's content without incurring any related-expense.

Nonetheless, *Meltwater* provides guidance to news aggregators, to other entities that use scraped content to generate analytics and to content-owners.

First, the *Meltwater* court recognized that, although increasing access to information has been recognized as a transformative use, and that "there is a strong public interest in preserving th[e] ... instantaneous[] and efficient access to information" that information-location tools such as search engines provide,³⁶ the use of copyrighted content for this purpose is not always a transformative. Accordingly, *Meltwater* suggests that news aggregators should consider avoiding reproducing substantial excerpts of copyrighted materials, to supplement any excerpted material with analysis or commentary and to structure their use of aggregated

³² *Id.* at *26.

³³ *Id.*

³⁴ *Id.*

³⁵ *Id.*

³⁶ *Id.* at *13.

content to minimize the likelihood that it would serve as a substitute for the excerpted source material.

Second, Meltwater establishes that a party cannot rely on the failure of a website to deploy the robots.txt protocol in order to avoid liability in connection with unauthorized web-crawling and scraping. The court's holding on this point benefits website owners, who otherwise would potentially face the choice between deterring welcome web crawlers such as search engines that might increase traffic to the website if the protocol were deployed, and surrendering control over copyrighted content on the website if the protocol were not. Nonetheless, because the court distinguished the circumstances before it from those in cases holding that a plaintiff is estopped from asserting copyright claims in connection with web-crawling that it knew would result in the absence of the protocol,³⁷ website owners should not interpret *Meltwater* as establishing a blanket rule that failure to deploy the protocol cannot be asserted as a defense in connection with unauthorized crawling or scraping.

Third, while the court did not directly address whether Meltwater's scraping of the AP's content for the purpose of generating analytics relating to news coverage and sources was transformative, since the AP did not allege any copyright violations in connection with Meltwater's analytics services, the court's discussion of Meltwater's use of the AP's content to generate analytics suggests that such use is transformative, and that such analytics, if not a separate service from Meltwater's excerpting service, would have increased the extent to which Meltwater's excerpting is transformative.³⁸ This suggestion is consistent with the Second Circuit's recognition that the use of copyrighted content as the "raw material" to "creat[e] ... new information, new aesthetics, new insights and understandings" is the "very type of activity that the fair use doctrine intends to protect for the enrichment of society."³⁹ Thus, the *Meltwater* decision may be viewed as suggesting implicitly that a copyright claim may not lie in connection with the unauthorized use of scraped content for purposes of data analytics under the right set of circumstances. (*Meltwater* does not contain any suggestion as to whether such a claim could succeed if based on other legal theories, such as breach of contract).

In conclusion, content owners, web aggregators and others who make use of content on the Internet should be cognizant of the *Meltwater* decision when considering the implications of posting and use of online content.

³⁷ See, *supra*, note 31.

³⁸ See *id.* at *17.

³⁹ *Blanch v. Koons*, 467 F.3d 244, 251-52 (2d Cir. 2006) (quoting *Castle Rock Entm't, Inc. v. Carol Publ'g Group, Inc.*, 150 F.3d 132, 142 (2d Cir. 1998)).

Circular 230 Disclosure: Internal Revenue Service regulations provide that, for the purpose of avoiding certain penalties under the Internal Revenue Code, taxpayers may rely only on opinions of counsel that meet specific requirements set forth in the regulations, including a requirement that such opinions contain extensive factual and legal discussion and analysis. Any tax advice that may be contained herein does not constitute an opinion that meets the requirements of the regulations. Any such tax advice therefore cannot be used, and was not intended or written to be used, for the purpose of avoiding any federal tax penalties that the Internal Revenue Service may attempt to impose.

Bingham McCutchen®

© 2013 Bingham McCutchen LLP

One Federal Street, Boston, MA 02110-1726

ATTORNEY ADVERTISING

To communicate with us regarding protection of your personal information or to subscribe or unsubscribe to some or all of our electronic and mail communications, notify our privacy administrator at privacyUS@bingham.com or privacyUK@bingham.com (privacy policy available at www.bingham.com/privacy.aspx). We can be reached by mail (ATT: Privacy Administrator) in the US at One Federal Street, Boston, MA 02110-1726 or at 41 Lothbury, London EC2R 7HF, UK, or at 866.749.3064 (US) or +08 (08) 234.4626 (international).

Bingham McCutchen (London) LLP, a Massachusetts limited liability partnership authorised and regulated by the Solicitors Regulation Authority (registered number: 00328388), is the legal entity which operates in the UK as Bingham. A list of the names of its partners and their qualification is open for inspection at the address above. All partners of Bingham McCutchen (London) LLP are either solicitors or registered foreign lawyers.

This communication is being circulated to Bingham McCutchen LLP's clients and friends. It is not intended to provide legal advice addressed to a particular situation. Prior results do not guarantee a similar outcome.