News aggregators and copyright in the European Union and the United States in the digital age: Evolution, comparisons, and implications

by Qun Wang and Susan Keith

Abstract
News aggregators have triggered copyright-related disputes between tech companies and news publishers. In the EU and the U.S., copyright systems have developed distinct characteristics. Because American tech companies stand to be hugely affected by the EU’s new copyright rules, some observers point out that the copyright war in Europe is fundamentally a collision between European and American copyright law systems. To respond to this observation, this study examines and compares European and U.S. perspectives on copyright and uses copyright as a lens to explore how digital platforms that aim at global influences provide the opportunity for different legal systems and legal traditions to converse and conflict. Through the comparison, this study argues that fundamental issues such as the nature of news are not effectively addressed in both systems. While the EU and the U.S. present different regulatory trends in the case of copyright, a two-way shaping is at play.

Contents
Introduction
Copyright law in the European Union
Copyright law in the United States
Conclusion

Introduction
News aggregators have become important actors in the twenty-first century news media ecosystem. In the digital context, a news aggregator is a Web application or a Web site that engages in the practice of news aggregation, which typically involves collecting news from various news sources, rearranging the collected news based on certain procedures, and distributing the rearranged news on its own Web site. There are different types of news aggregators. For example, based on the types of sources used, there are feed aggregators, sites that organize news into “feeds” by topic or story; specialty aggregators, Web sites that collect information from various sources on a particular topic; user-curated aggregators, which aggregate headlines and articles shared by the public; and blog aggregators, which use third-party material to compile blog entries (Isbell, 2010). In terms of the approaches on which their operations are based, some news aggregators primarily depend on automated computational algorithms, others apply human judgment, and still others use a mix of both (Lee and Chyi, 2015).

Large tech companies have news aggregation services, such as Google’s Google News, Apple’s Apple News, Microsoft’s Bing News, and Facebook’s Facebook News. Traditional news media have also launched their own news aggregators. News Corp. started Knewz in early 2020. CNN was working on an aggregator known internally as NewsCo in 2019 and 2020 (Coster, 2020; Lunden, 2020; Toonkel, 2019). Globally, Reuters Institute’s 2020 Digital News Report, based on a survey of more than 80,000 people in 40 countries that cover Europe, Asia, Latin America, North America, and Africa, found that news aggregators are an important gateway through which users get news, especially on mobile platforms (Newman, et al., 2020).
While they provide an efficient way for users to access news, aggregators are highly controversial because of the way they handle news and their relationship with the news media. Typically, news aggregators do not produce news themselves, but they play the role of gatekeeper often attributed to traditional news outlets as they influence what news users are exposed to and in what ways they see that news. The rise of news aggregators and their practices in using different ways to select, rank, and present news have produced a great deal of controversy. In March 2010, at a workshop held by the U.S. Federal Communications Commission, news aggregators associated with search engines were named the “enemy” of the news industry (Anderson, 2013). Google News, the world’s largest news aggregator, has been involved in a series of disputes worldwide (Wang, 2020). A major concern of news publishers involved in these disputes has been copyright.

Copyright is a set of legal rights granted, in modern times, to the authors of original expressive works (Netanel, 2018). It was intended, as Baldwin (2014) writes, “to give authors sufficient encouragement to remain fruitful, thus enriching the public domain and serving useful social functions — to enlighten, entertain, and educate” [1]. Although varied historiographies of copyright have arisen in different centuries (Deazley, et al., 2010; Maxwell, 2004), recognition of legal rights to control new expressive works often has been tied to the invention and political economy of more efficient technologies for copying those works (Bettig, 1996), such as the Internet.

One example emerged in early 2021 when the Australian Parliament passed the News Media and Digital Platforms Mandatory Bargaining Code (Parliament of Australia, 2021). This law requires that platforms, including those such as Google that have aggregators, compensate certain publishers of “core news content” for online use of their content. The mere threat of the law requiring platforms such as Google and Facebook to pay local publishers for their content (Lomas, 2021), prompted Google to work out payment deals with more than 70 Australian titles. It led, however, to Facebook banning news sharing by Australian Facebook users before reversing that decision and also making deals with news companies (Morrison, 2021).

Historically, however, the debate between local publishers and platforms has been especially intense in the 27-nation European Union, which in 2019 approved a new copyright directive (Directive (EU) 2019/790, 2019). Because American tech companies stand to be hugely affected by the new European rules, which must be incorporated into the laws of EU member states by June 2021, some observers point out that the copyright war in Europe against large news aggregators is fundamentally a collision between European and American copyright law systems. For these reasons, this study focuses on the European and U.S. contexts, providing an in-depth interpretive legal analysis of their perspectives on copyright. It also uses copyright as a lens to explore how digital platforms that aim for global influence provide the opportunity for different legal systems and legal traditions to converge and conflict.

This study relies on interpretive legal analysis of the type addressed by Milosavljević and Poler (2019), who point out that for macro analysis focusing on legal situations at a national or supranational level, a legal comparison might be the “most effective solution” [2]. Following steps identified by Milosavljević and Poler, we gathered and examined relevant cases, particularly copyright cases related to news aggregators, traditional news media, and tech companies. We also reviewed and examined legal and non-legal materials related to the discussion — drawing on wider legal frameworks of copyright in both the U.S. and the EU, fair use, the hot news doctrine, and Digital Millennium Copyright Act — to understand procedures, precedents and doctrine, as well as their developments, interpretations, and their uses within different environments and over time.

Copyright law in the European Union

Background

Contemporary European copyright law developed from the Berne Convention, an international intellectual property treaty signed in Berne, Switzerland, in 1886, which aimed to provide creators “with the means to control how their [the creators’] works are used, by whom, and on what terms” (World Intellectual Property Organization [WIPO], 1979). As this treaty was drafted to address copyright issues in the pre-Internet era and was last updated in 1971, newer copyright systems were required to deal with new developments in the digital environment. In 1996, the World International Property Organization issued international copyright standards as a response to this need. The WIPO Copyright Treaty — concluded in 1996 and entered into force in 2002 — adopted principles of the Berne Convention but also put forward a digital agenda to address subject matter to be protected by copyright in the digital environment, such as computer programs and databases, that were not covered by the Berne Convention (World Intellectual Property Organization [WIPO], 1996). The WIPO required its member countries to update their copyright laws to comply with the WIPO treaties (Blythe, 2006). It was against this backdrop that a series of legislative proposals were put forward aiming at the modernization of EU copyright rules in the information society.

While these proposals attempted to balance the interests of right holders and other actors in the information society, such as network operators, the EU Copyright Directive (EUCD) that went into effect in 2001, in which the essential parts of the current EU copyright rules are rooted, tended to refocus on high level protection of rights-holders and fair compensation. It defined a set of rights, such as the reproduction right, the right of communication to the public, the distribution right, and moral rights,
and offered relatively limited exceptions and limitations. Article 5(5) of the EUCD 2001 noted that the exceptions and limitations shall only be applied in “certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder” (EUCD, 2001).

The goal of the EUCD was to avoid fragmented legal approaches implemented by national provisions in order to ensure through harmonized copyright rules the proper functioning of an internal market in the EU and the competitiveness of European industry. The EUCD concerned “intermediaries” or “transmission systems” as one of the parties involved in the copyright system in the information society. It was not until 2019, however, that news aggregators became an addressee of the new EUCD at the EU level. In those years between 2001–2019, news aggregators, especially Google News, were involved in a number of lawsuits filed by European news publishers. In the aftermath of these legal disputes, some EU countries passed legislation that particularly addressed news aggregators and their copyright obligations.

**Copyright lawsuits against news aggregators and national provisions in the EU**

In the EU, the tension between news publishers and news aggregators seems to have escalated since Google’s news aggregation service entered the European market in 2003. Due to the lack of copyright rules that specifically address news aggregators and their practices, much of the early litigation between news publishers and news aggregators concluded with settlements. For example, in 2005, Agence France Presse (AFP) sued Google News for copyright infringement, but the two parties reached an undisclosed licensing agreement in 2007 (Auchard, 2007; Keith, 2007). The copyright lawsuit between Belgian publishers’ association Copiepresse and Google in 2006 was also settled with an agreement in 2012 (Essers, 2012). In other EU countries, such as Portugal, Austria, and Liechtenstein, national laws tended to grant safe harbors for search engines, but provisions particularly addressing news aggregators were absent for many years (Xalabarder, 2012).

In 2013, Germany passed the Ancillary Copyright Law for Press Publishers, which was one of the earliest efforts in the EU to regulate news aggregators through new legislation. The law granted German news publishers the exclusive right to their press products and stipulated that search engines and news aggregators shall pay “equitable remuneration” to use the publishers’ press products (BGBl, I S. 3728, 2013). In June 2014, VG Media, the collective that represented over 230 German publishers, sued Google for copyright violation under the Ancillary Copyright Law, but Google was able to bypass the compensation liability by only displaying the hyperlink and headline to articles from German publishers that did not opt in, as “individual words or very short text excerpts” were permissible under the law. Shortly after, VG Media and associated German publishers gave Google permission to host their content for free due to the “major economic pressure” of losing traffic (Hebbard, 2014).

Having learned lessons from Germany and early cases in other EU countries, AEDE, the Association of Spanish Daily Publishers, lobbied for a stricter copyright law in Spain. In 2014, Spain updated its intellectual property law, aka Canon AEDE, which specifically defined news aggregators’ copyright obligations. Except for the use of non-significant fragments of the press content, the Canon AEDE required an unwaivable right for fair compensation for Spanish news publishers when news aggregators used their content. Search engines, however, were excluded from such obligations (Boletín Oficial del Estado, 2014). As a result, Google News shut down its service in Spain before the official implementation of the law. Two major features of the Canon AEDE — the mandatory nature of compensation and the distinguishing of news aggregators from search engines — distinguishes Spain’s copyright law from other EU countries. While the requirement of compensation resulted in the shutdown of Google News in Spain, the discrimination between news aggregator and search engine gave Google search the opportunity to continue to host Spanish publishers’ news content.

**The new EUCD**

As different countries adopted different approaches and national provisions, which resulted in varying outcomes, the EU was concerned about the refragmentation of the internal market and legislative inconsistency. After struggles and compromises among the “trilogue” — the European Parliament (the legislative branch), the European Commission (the executive branch), and the Council of Europe (a broader group of 47 European countries, including 20 not in the EU) — on 26 March 2019, the European Parliament approved the new Copyright Directive, which aimed at a “harmonized legal protection for press publications in respect of online uses by information society service providers” (European Parliament, 2019). The Directive entered into force on 7 June 2019, and EU member states were given the deadline of June 2021 to implement the new copyright rules into their national legislation (Meyer, 2019).

The new EUCD contains two particularly controversial provisions: Article 11 and Article 13 of the draft version, which became articles 15 and 17, respectively, of the final directive. Critics have dubbed the two articles a “link tax” and an “upload filter” (Vincent, 2019). Most relevant to this study is Article 15 (formerly Article 11), which grants protection of press publications concerning online uses and allows EU press publishers the right to claim revenue from online use of their publications. The new EUCD explicitly identifies news aggregators as part of the information society service providers in the digital environment that the Directive aims to address; but the legislation exempts acts of hyperlinking and the use of “individual words or very short extracts of a press publication” (Directive (EU) 2019/790, 2019).

**What has been addressed in the new EUCD**
The new EUCD has practically responded to earlier copyright lawsuits against news aggregators by adopting or changing certain legislative approaches in national laws of its member states. At the EU level, the new EUCD grants a new neighboring right for press publications — “journalistic publications, published in any media ... for instance, daily newspapers, weekly or monthly magazines of general interest, including subscription-based magazines, and news websites” — to authorize or prohibit the online uses of their content by digital platforms, as well as the right for fair remuneration for press publishers and individual journalists. Under the new EUCD, news professionals have joined other creators, such as musicians and actors, in having their news works taken into the copyright framework in the digital time. While the new EUCD adopts the news publishers’ entitlement to compensation, the mandatory nature is not as strong as that of Spain’s Canon AEDE and it does not explicitly discriminate between search engines and news aggregators as Germany’s ancillary copyright law does.

This is a move toward an increased degree of regulation over information society service providers, including news aggregators. Using “information society service providers” as a general addressee blurred the boundary between search engine and news aggregator, which could help prevent online service providers from circumventing legal regulation due to the discrimination, as happened in the Spanish case. This change indicates the recognition of the complex nature and connection of tech companies and their practices. For example, when Google attempted to avoid copyright liability by defending its news aggregation service as a search engine service, the court ruled that Google News is an “information portal” functioning beyond the role of a search engine because it can algorithmically manipulate information (Copiepresse v. Google Inc., 2011). The indiscriminate approach also makes the new EUCD almost exclusively target Google — an American company that has both search engine and news aggregation services — and therefore confirms the “Google tax” label that observers attached to related legislative developments in EU’s copyright systems and leaves the EU not completely innocent of the charge of being protectionist.

News aggregators commonly make a few news elements available on their platforms, including headlines, visuals, snippets, and hyperlinks. In the EU, news publishers argued that the combination of these key elements of online news greatly reduced the likelihood that users would go to and browse the news Web sites that originally provide these news stories (Agence France Press v. Google Inc, 2005). To address the use of these elements, the new EUCD explicitly excludes the use of hyperlinks from copyright obligation but protects photographs and videos for press publications. It exempts the use of “individual words or very short extracts” of press publications but does not specify the exact number of words that are permitted for online platforms to use without authorization (European Parliament, 2019). Such statutory terms could affect both news headlines and news snippets — as they both contain individual words and short extracts of news content — leaving room for the EU’s member states to interpret this factor differently in their actual implementations.

Under the new EUCD, licensing is used as the main mechanism for press publishers to control the online uses of their works and obtain remuneration for such use. In the EU, since government subsidies are generally not adequate to financially sustain news media in the digital era (Trappel, 2018), licensing could be an alternative in the EU context. The licensing mechanism would also allow publishers to set the conditions on their own choices. For example, small publishers who are willing to opt in news aggregators’ services may grant free licenses to maintain their online visibility. Such benefits will only be achieved through a stronger negotiation power. In the EU, any individual country alone may not be able to compete with powerful companies like Google and Facebook, but the EU as a united market helps form stronger regulatory and economic pressure for these tech giants to compromise, as shown in Google’s recent licensing program to pay publishers in Europe and other parts of the world (Barker, 2020). In this regard, the harmonized copyright protection is more of a utilitarian means to maintain the functioning of the internal market and the competitiveness of EU industries. After all, the copyright-intensive sectors of the EU economy account for over seven million jobs, or three percent of employment, in the EU (European Commission, 2019). From this perspective, the new EUCD is a move to benefit the EU’s economic interest.

And what remains unsolved

The provisions discussed above tend to offer practical measures to respond to short-term controversies involved in early copyright disputes between EU publishers and news aggregators, many fundamental issues, however, have not been effectively addressed in the new EUCD.

First, the nature of news. News was not traditionally protected under the copyright law since copyright law concerned the literary, scientific and artistic domains and the creative works and innovations (World Intellectual Property Organization [WIPO], 1979). News, on the other hand, has been in the domain of public service to inform the public. Article 2(8) of the Berne Convention addressed that the copyright protection shall not apply to news and facts in press information (World Intellectual Property Organization [WIPO], 1979). When defining rightholders, Article 2 of the old version of EUCD in 2001 did not include press publishers either. The new EUCD changed this tradition by incorporating news into copyright protection, but it also points out that the rights granted to press publishers should not extend to the “mere facts” reported in the press publications” (2019, emphasis added). Instead of clarifying the nature of news, the new EUCD complicates it and leaves previous debates, such as the one about the copyrightability of news in AFP v. Google, unsolved (Keith, 2007). Where is the line between facts and non-facts in news? Is news privately owned intellectual property or public service that the public can have free access? Answers to these questions may vary in different contexts, for example, French journalism has a political-literary style, while American journalism embraces fact-based reporting (Benson, 2013). The new EUCD reflects the
News aggregators and copyright in the European Union and the United States in the digital age: Evolution, comparisons, and implications

Second, by imposing copyright obligations on online service providers, the new EUCD has adopted an opt-in model, which is a departure from the opt-out trend in the Berne Convention and the EUCD in 2001. The opt-in model differs from the opt-out model in the way that the former assumes that content cannot be used without authorization unless publishers explicitly opt into the given online service, while the opt-in model assumes the opposite: that content can be used without authorization unless publishers explicitly opt out of the given online service. Article 7 of the original Berne Convention held that news articles could be reproduced “unless the author or editor has expressly prohibited so” (original text in World Intellectual Property Organization [WIPO], 1986), clearly an opt-out approach. Article 5 (3) of the 2001 EUCD also allowed certain exceptions or limitations to the granted copyright if the use is “not expressly reserved.” Under those provisions, online service providers had to remove the infringing content if the author or publisher clearly chose to opt out (2001/29/EC, European Union, 2001). In the opt-out model, the burden was placed on authors and publishers, who had to clearly express their opt-out request. Under the new EUCD, however, the burden is placed on news aggregators, who will have to make effort to obtain licenses from publishers and can use significant news content only if publishers are willing to opt in their online services.

Which model should be adopted is an important question as it’s a matter of who has the control. While a shift from an opt-out model to opt-in model is enabled by the passage of the new EUCD, the shift is not effectively justified in the new copyright law. In a few lawsuits filed by EU publishers against Google, for example, Google defended the opt-out model by arguing that industry standards, such as meta tags and robot exclusion protocols, are standard technical tools for publishers to use to inform online crawlers whether publishers want their content to be crawled. Without using these tools, Google argued, publishers did not choose the opt-out option while this option is available. In the new EUCD, these technical issues and their implications for control and the power relations between platforms and news media were not addressed while technology-oriented issues should not and cannot be bypassed when law and policy-makers make regulatory decisions over technology companies.

---

Copyright law in the United States

Background

The United States, like many EU countries, is a member of both the Berne Convention (since only 1988) and the WIPO, but there have long been notable differences between the U.S. view of copyright and European view. The U.S. Constitution, ratified in 1787, gave Congress the power to “to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries” (U.S. Senate, n.d.). This phraseology implicitly articulated an economic foundation for copyright, rooted in the Enlightenment-era utilitarian philosophical tradition (Jacobs, 2016), rather than the “droit moral,” or moral-rights-of-the-author tradition more prevalent in Europe (Kikkis, 2012). The first U.S. federal copyright law, passed in 1790, gave U.S. citizens, residents, or their executors or administrators the sole right to print, reprint, publish, or sell copyrighted works for a period of 14 years (U.S. Copyright Act of 1790, U.S. Copyright Office, n.d.a.), a term that has been lengthened over time to the present term of life of the author plus 70 years. The last full update to U.S. copyright law occurred in 1976 (U.S. Copyright Act of 1976, U.S. Copyright Office, n.d.b.), though Congress amended several portions of that law with the Digital Millennium Copyright Act of 1998.

Unlike the EU, the United States has not enacted a law dealing specifically with news aggregation, and only one federal court has ruled on whether aggregator practices are legal — in a case that was appealed and then settled out of court (Associated Press v. Meltwater Holdings Inc., 931 F. Supp. 2d 537, 2013). Various common law and statutory approaches are relevant, however, to how the country has approached practices that involved aggregation. This section focuses on three: the fair use doctrine of U.S. copyright law, the “hot news” doctrine applied in a key early twentieth century U.S. Supreme Court case, and the Digital Millennium Copyright Act’s prohibition of technologies designed to circumvent copyright.

U.S. legal frameworks relevant to aggregation

Fair use

The U.S. doctrine of fair use makes legal certain limited uses of copyrighted works without the copyright holder’s permission. It often is traced to the nineteenth century case Folsom v. Marsh (C.C.D. Mass., 1841), though Sag (2011) argues that its roots lie in the doctrine of “fair abridgement” developed in pre-colonial English courts. In Folsom, Massachusetts Judge Joseph Story outlined the four factors that courts have since used to assess whether use of a copyrighted work was permissible “fair use”: 1) the nature of the use, including whether it is for non-profit or educational use; 2) the nature of the copyrighted work, including whether it was expressive or factual and whether publication had already occurred; 3) the amount and substantiality of the original work that was used; and, 4) the effect on the potential market for the original work. Congress codified this common law doctrine into statutory law in its 1976 revision of the Copyright Act.
In 1994, the U.S. Supreme Court added to the first fair use factor a new twist: the doctrine of transformative use, which holds that the more an unauthorized use of a copyrighted work changes the original, the more likely the use is to be fair use. For example, parody was ruled to be fair use in *Campbell v. Acuff-Rose Music, Inc.* Although some have criticized what they see as the subsequent over-use of the fair use doctrine (Bunker, 2002), its use in search engine-related cases from the U.S. Ninth Circuit Court of Appeals — including *Kelly v. Ariba Soft* (336 F.3d 811, 2003) and *Perfect 10, Inc. v. Amazon.com, Inc.* (508 F.3d 1146, 2007) — suggests that the categorization and indexing functions of feed aggregators might make their unauthorized use of news adequately transformative to constitute fair use. User-curated and blog aggregators have an even stronger case, Isbell (2010) has argued, as their uses of news can be so transformative that the aggregators no longer reproduce the original but imbue the content with a new meaning or message.

Some scholars have argued that news aggregators would be able to claim fair use on the second factor, the nature of the work that was used, for at least two reasons. First, the news articles compiled by aggregators are generally based in facts, which cannot be copyrighted under U.S. law. Second, the articles have been published, a factor that weighs in favor of the unauthorized user, before they are aggregated (Isbell, 2010). The third factor — the amount and substantiality of the portion of the copyrighted work that was used — is debatable. On the one hand, many news aggregators use only the headline of a news article and perhaps a lead or summary paragraph; on the other hand, those portions of a news article could be considered the “heart of the work.” Use of the heart of a work, even if it is only a small portion of the whole, could be the basis for a successful copyright claim, the U.S. Supreme Court ruled in *Harper & Row v. Nation Enterprises* (471 U.S. 539, 1985).

It is the fourth factor of fair use — the effect of news aggregators on the market for original news stories — where the chief tensions could lie between U.S. news producers, which have generally characterized news articles as valuable, expensive-to-produce works, and aggregators seeking to distribute information widely. Scholars such as Xalabarder (2012) have maintained that while the fourth element could be the most important element of any fair use analysis, its judgment depends on various factors (Xalabarder, 2012). For example, the Associated Press accused Google’s news aggregation of harming the AP’s business model and insisted that sites using AP content should share advertising revenues with the wire service, whereas Google argued that its news aggregation service is ad-free and it could benefit news sites by sending them online traffic (Pérez-Peña, 2009; Claburn, 2007).

In the only U.S. copyright infringement case involving a news aggregator, the defendant attempted, ultimately unsuccessfully, to use the fair use defense. In *Associated Press v. Meltwater Holdings Inc.* (2013), the Associated Press argued in the U.S. District Court for the Southern District of New York that the Norwegian company Meltwater Holdings Inc. had violated the wire service’s copyright by providing a private, subscription-based service that used computer programs to aggregate and display portions of AP news articles that mentioned Meltwater’s clients. Although the court conceded that the second factor, nature of the copyrighted work, was neutral, Meltwater failed to gain support from the court on factors 1, 3, and 4. The court ruled that Meltwater had violated the AP’s copyright. Nevertheless, the ruling established a precedent, in a court based in the U.S. media capital, for viewing the role of aggregators as distinct from that of search engines and thus more vulnerable to copyright infringement and “heart of the work” claims. The implication of the Meltwater case is debatable in the U.S. context as Meltwater’s ownership, scale, and influence are hardly comparable with that of large U.S. tech companies. But overall, Xalabarder (2012) argued, use of the fair use doctrine may not ensure a win in a copyright lawsuit by either a news aggregator or a news media company.

**Hot news doctrine**

Another legal framework relevant to the legality of news aggregators in the U.S. is hot news misappropriation. This state-law tort sprang out of the 1918 U.S. Supreme Court decision in *International News Service v. Associated Press* (248 U.S. 215), a case in which two U.S. news agencies battled over whether compilations of facts in the form of news stories could be copyrighted. In the lawsuit, the AP, a collective of about 950 member newspapers, charged that INS had “paid off AP-member telegraphers to secure early editions of AP news, then recast that news for its own subscribers” about 400 newspapers [3]. INS argued that once the Associated Press’ output was published, “the right of property is lost, and the subsequent use of the news ... for any purpose whatever becomes lawful” [3]. Writing for the 5-3 majority, Associate Justice Mahlon Pitney defined a “quasi-property right” in breaking news because the “peculiar value of news is in the spreading of it while it is fresh” [3]. Based on this “hot news” doctrine, the court held that INS had engaged in misappropriation, selling AP’s articles “as its own” [3].

Scholars such as Applegate and Schermerborn (2011) have pointed out that although the Court held that news facts were not copyrightable as they were the “history of the day,” it also acknowledged the cost, skill, labor, and money that news media invested in generating news, especially the fresh, time-sensitive hot news. The hot news doctrine, the authors maintained, is an effort to “provide an incentive” for news publishers and other parties to “promote the growth of knowledge,” the goal of the Copyright Clause [8].

In 1997, the U.S. Second Circuit Court of Appeals set forth a five-part test to determine hot news misappropriation. The central elements are: (i) the plaintiff generates or collects information at some cost or expense; (ii) the value of the information is highly time-sensitive; (iii) the defendant’s use of the information constitutes freeriding on the plaintiff’s costly efforts to generate or collect it; (iv) the defendant’s use of the information is in direct competition with a product or service offered by the
plaintiff; (v) the ability of other parties to free-ride on the efforts of the plaintiff would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened [9]. The five-part test was argued to be more relevant in the digital age as it was “considered with modern technology such as Google, Twitter, Flickr, and other Internet-based modes of information dissemination in mind” (Kerschberg, 2011).

The five-part test, however, adds uncertainties for news publishers that attempt to pursue the hot news doctrine in copyright infringement lawsuits, as news aggregators are likely to defend themselves from most of these elements. They could argue that (i) they aggregate both time-sensitive and non-time-sensitive news; (ii) that their practice is not free-riding because they have their own technological and intellectual investment; (iii) that their news aggregation service is not in direct competition with news media as they are different in many ways; and (iv) that their news aggregation service sends online traffic to news publishers and therefore would not reduce the incentive of news publishers. Jurisdiction could play a role, too, as only a minority of states recognize hot news misappropriation as a tort, though, importantly, the media center of New York is one of them.

Digital Millennium Copyright Act

Like many European countries, the United States is a member of the WIPO. In response to the WIPO requirement that its member countries update their copyright law for the digital era, Congress passed the Digital Millennium Copyright Act (DMCA) in 1998. That act, embedded in the American economic/utilitarian view of copyright and heavily influenced by interests from what are sometimes referred to as the “copyright industries” (Imfeld and Ekstrand, 2005), made it illegal to circumvent technological measures used by copyright holders to prevent copying of digital and electronic works. In spite of the DMCA’s digital focus, the technological issues involved in the rapidly changing media and technology environment are often more complicated than the DMCA could address, especially when those large news aggregators are also leading technology companies that have strong capability in technology innovation.

For example, some scholars have argued that a robot.txt file — a file that tells Internet crawlers, spiders, or robots which files they may not access — does not make it technologically impossible to aggregate information from a site that uses the files (Shipley and Bowker, 2013). If a news aggregator ignore warnings from robot exclusion files, it could, at least theoretically, be seen as using technology used to circumvent digital rights management protections in violation of the DMCA. Interestingly, in the Meltwater case, the defendant aggregator tried to argue that because the Associated Press does not direct sites that license use of its news to use robot exclusion protocols, the wire service was granting Meltwater an implied license to use AP news. The federal district court rejected this argument on multiple grounds, including the fact that it 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” [10].

Meanwhile, the DMCA offers safe harbors for Internet companies in several categories, including companies that provide Internet connectivity, transmitting, routing; companies that provide caching; companies that provide information location tools, including indexes, hyperlinks, and directories; and companies that host or store material on the Internet at the behest of their users (Travis, 2008). The DMCA safe harbors have important implications for news aggregators in the U.S. because technologies protected by DMCA safe harbors are commonly used in digital news aggregation services.

Characteristics of the U.S. copyright system

Many copyright scholars maintain that the significance of the U.S. copyright framework lies in two dimensions. Economically, the U.S. copyright law takes a utilitarian approach to ensure the incentive for the progress of science, arts, and new knowledge, as outlined in Article I, section 8 of the U.S. Constitution (Hugenholtz, 2001). Socially, copyright also aims to further the First Amendment value in encouraging free expression and promoting the access to information of the informed public (Netanel, 2018). Some scholars argue that the two considerations are not separable, as fiscal independence is necessary to sustain free expression (Netanel, 2018). In the U.S. copyright system, the pendulum has swung back and forth in order to maintain the nuanced balance between the two ends when weighing the interests of copyright owners and users; courts and Congress have shifted from “trade control for compensation” and “some compensation some control” models (Ginsburg, 2001).

New developments in the digital age have raised new challenges to the U.S. copyright system. As far as large digital news aggregators are concerned, one pattern that distinguishes the U.S. from the EU is that U.S. news publishers, unlike their European counterparts, have not tended to pursue litigation against news aggregators, although conflicts have existed (Author). When lawsuits were filed, disputes were often settled through contractual agreements. For example, there was a months-long impasse between the Associated Press and Google in 2009–2010 regarding how Google could use the AP’s news content. Google stopped hosting the AP’s news until the two parties reached a new agreement in August 2010 (Cohen, 2010). Like the AP, very few U.S. news organizations seek legal actions when dealing with their issues with news aggregators in the digital age. There are a few reasons that contribute to this different pattern between the U.S. and European news publishers when dealing with the copyright issue with news aggregators.

The nature of news
Is news copyrightable? This question remains largely unresolved in both the U.S. and EU copyright systems. Since the U.S., like the EU, is a member of the Berne Union, it traditionally has limited copyright protection over news articles, as addressed in the Berne Convention 1886 that news articles could be reproduced “unless authors or editors had expressly reserved so” (Berne Convention, Article 7 as of 1886; see World Intellectual Property Organization [WIPO], 1986). Certain elements of news, such as headlines and short phrases are also traditionally not protected in the U.S. (Xalabarder, 2012). However, the copyrightability of news is contested in the U.S. context, given that the facts/non-facts dichotomy has greater meaning in the United States, as its journalistic tradition values fact-based news. Article 2 (8) of the Berne Convention stipulates protection shall not apply to “news of the day or to miscellaneous facts having the character of mere items of press information” (World Intellectual Property Organization [WIPO], 1979). Although there are different interpretations, Article 2 (8) seems to point to a facts/non-facts dichotomy. However, the time-sensitiveness of news facts is protected under the hot news misappropriation doctrine in the United States, at least in jurisdictions that recognize the tort. While news aggregators handle news based on both facts and highly expressive content, these legal frameworks, instead of clarifying, complicate the copyrightability of news.

The regulatory trend

Under the fair use provision defined in section 107 U.S. Copyright Act and the safeharbor protections granted in the DMCA, internet companies are more likely to enjoy copyright exemptions. In the U.S., some technology companies successfully defended lawsuits for copyright infringement, in which courts ruled that these companies’ practices, such as indexing images, caching, etc., practices that are also common for news aggregators, were fair use (e.g., Field v. Google, Inc., 2006; Kelly v. Arriba Soft Corp, 2003). News aggregators, Xalabarder (2012) has argued, are likely to make successful defense in at least three of the four fair use factors. While it is “impossible to forecast a univocal solution to whether news aggregation as a whole is fair use” [11], it’s equally true whether it’s not fair use. Such uncertainties add risk to news publishers that seek legal actions under the copyright framework in the U.S. Overall, although there are shifting trends in the U.S. copyright regulation, courts and Congress have showed solicitude for new markets (Ginsburg, 2001), as “the adoption of new technology that injures or destroys present business models is commonplace,” the U.S. Second Circuit Court of Appeals ruled in TheFlyOnTheWall.com case, “Whether fair or not, that cannot, without more, be prevented by application of the misappropriation tort” (Barclays Capital Inc. v. TheFlyOnTheWall.com, 2011).

Conclusion

In the digital information landscape, news aggregators have a profound impact on news media and democracy. Their rise and the corresponding digital issues have also required reconsiderations of current copyright frameworks in the European Union and the U.S. contexts. This study provides reflections on the comparison of the two copyright systems, which largely shape the two areas’ media and technology development. In the digital era, both the EU and the U.S. updated their copyright regulations to respond to the WIPO’s requirements. Their respective copyright systems, however, have been developing along different paths. Compared to the U.S., where the copyright issue between news publishers and news aggregators is mostly an internal issue, the EU copyright system is facing a more complex internal and external environment. Large news aggregators involved in copyright disputes in the EU are often U.S. tech companies. When dealing with these disputes, the EU’s role as the “pan-European copyright legislator” is at play for it has to consider the EU’s internal market and the European consensus (Hugenholtz, 2001).

There has been a two-way shaping between the EU and the U.S. copyright systems. In the EU context, given the economic pressure, despite the legal tradition that treated copyright as moral right, the EU’s recent copyright reforms adopted the U.S. utilitarian approach to address immediate, economic concerns of the news industry. But compared to the U.S. copyright system, in which the regulatory pendulum has been swung in order to maintain the balance between economic incentive as well as First Amendment values, in the EU, it is “improbable the European Court, in light of its deference to consensus, will be easily convinced to apply Article 10 in order to restore the equilibrium” [12]. After all, although freedom of information is acknowledged in the EU, U.S. tech companies are not considered the main bearers of such values in Europe.

Meanwhile, recent legal developments in Europe have informed the regulatory climate over digital platforms in the U.S. For example, Google recently announced it will pay publishers to license content in France, Australia and other parts of the world, and Facebook is willing to pay about 200 U.S. publishers to use their content for Facebook’s News tab, the Facebook version of news aggregation service (Alpert and Patel, 2019). U.S. Congress has also been investigating dominant tech companies’ impact on traditional media industry, but the focus is on monopoly, indicating a more structural concern than particular economic right like copyright. At the same time, digital news aggregators in particular and digital platforms in general — many of them based in the U.S. but aimed at a global information market — have to respond to legal influences inside and outside the U.S. As the global information landscape is getting more complex, the U.S. legal system needs to take international influences into its regulatory consideration.

Besides these distinctions, in both the U.S. and Europe, the nature of news is not convincingly addressed, which is one of the
fundamental issues that would affect the philosophy and the effectiveness of the copyright as well as other legal frameworks when dealing with issues related to news media. In addition, technological issues were addressed differently in the U.S. and European copyright systems, suggesting the incompatibility between traditional legal logics and technological logics, a gap that needs to be filled through the collaboration of expertise from both areas. But technological issues often have implications beyond the technological domain and copyright is only a mirror that reflects the complex dynamics in the global information environment in the digital age.

About the authors

**Qun Wang**, Ph.D., is an assistant professor in the Department of Communication and Media Studies at Fordham University. A former journalist, Dr. Wang studies journalism, technology, and platforms. Dr. Wang’s research examines the interrelationship between news industry and tech industry, the news media’s digital transformation, and the dynamics in the global media and technology landscape. Her work has been published in such journals as *Journalism & Mass Communication Quarterly, Digital Journalism, Journalism Practice, Journalism: Theory, Practice and Criticism*, and *Journal of International Communication*. E-mail: qwang184 [at] fordham [dot] edu

**Susan Keith**, Ph.D., is an associate professor in the School of Communication and Information at Rutgers University and president elect of AEJMC. A former journalist, Dr. Keith studies the evolving practice of journalism, journalistic content (especially visual content and coverage of war and conflict), and legal and ethical concerns related to journalism. Her work has been published in such journals as *Journalism & Mass Communication Quarterly; Journalism: Theory, Practice and Criticism; Journalism Studies; Journal of Mass Media Ethics, Journal of Broadcasting & Electronic Media; Media, War & Conflict; and Visual Communication Quarterly*. E-mail: susank [at] comminfo [dot] rutgers [dot] edu

Notes

3. EUCD, paragraph 56.

References


V.S. Ekstrand and C. Roush, 2017. “From hot news to hot data: The rise of fintech, the ownership of big data, and the future of


News aggregators and copyright in the European Union and the United States in the digital age: Evolution, comparisons, and implications


News aggregators and copyright in the European Union and the United States in the digital age: Evolution, comparisons, and implications


Editorial history

Received 5 April 2021; accepted 17 August 2021.

Copyright © 2021, Qun Wang and Susan Keith. All Rights Reserved.