Big Data and the Fourth Estate: Protecting the Development of News Media Monitoring Databases

Joseph A. Tomain

Follow this and additional works at: http://digitalcommons.law.umaryland.edu/jbtl

Recommended Citation
Available at: http://digitalcommons.law.umaryland.edu/jbtl/vol12/iss1/4
Big Data and the Fourth Estate: Protecting the Development of News Media Monitoring Databases

I. INTRODUCTION

“Big data” is revolutionizing our lives and society. While big data fuels our information economy, it also raises privacy concerns in a variety of contexts including, but not limited to, political, consumer, student, law enforcement and others. There are various definitions for “Big Data,” including both technical and social definitions. Neil M. Richards & Jonathan H. King, Big Data Ethics, 49 WAKE FOREST L. REV. 393, 394 (2014) (providing several definitions and distinguishing between “big data” and “big data analytics”). A basic definition used here is: “[B]ig data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more.” VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 6 (2013).

1. There are various definitions for “Big Data,” including both technical and social definitions. Neil M. Richards & Jonathan H. King, Big Data Ethics, 49 WAKE FOREST L. REV. 393, 394 (2014) (providing several definitions and distinguishing between “big data” and “big data analytics”). A basic definition used here is: “[B]ig data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more.” VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 6 (2013).

2. President’s Council of Advisors on Science and Technology, Report to the President, Big Data & Privacy: A Technological Perspective i, x (May 2014) (“The beneficial uses of near-ubiquitous data collection are large, and they fuel an increasingly important set of economic activities.”); Susan Lund et al., McKinsey Glob. Inst., Game Changers: Five Opportunities for U.S. Growth and Renewal 73–74 (July 2013) (estimating nearly $610 billion in annual productivity gains and cost savings from widespread use of big data and analytics).

3. Ira S. Rubinstein, Voter Privacy in the Age of Big Data, 2014 WIS. L. REV. 861, 861 (2014) (proposing that greater transparency and new privacy restrictions on commercial data brokers may help to curb the potential privacy harms of voter microtargeting); Alex Emmons, Microsoft Pitches Technology That Can Read Facial Expressions at Political Rallies, THEINTERCEPT.COM (Aug. 4, 2016, 12:15 pm), https://theintercept.com/2016/08/04/microsoft-pitches-technology-that-can-read-facial-expressions-at-political-rallies/ (explaining that “political campaigns could use the technology to measure the emotional impact of different talking points and political scientists could use it to study crowd response at rallies”).

4. Andrew W. Bagley & Justin S. Brown, Limited Consumer Privacy Protections Against the Layers of Big Data, 31 SANTA CLARA HIGH TECH. L.J. 483, 484–85 (2015) (explaining that consumers give consent to online service providers, through opt-in and opt-out contracts, to share consumer data with third parties); Edith Ramirez, The Privacy Challenges of Big Data: A View from the Lifeguard’s Chair, Speech at Technology Policy
BIG DATA AND THE FOURTH ESTATE

national security.° Big data proponents commonly state that privacy concerns should not warrant legal restrictions that prohibit the development of new tools and uses of big data because regulation could stunt the growth of technology and the information economy itself, especially when actual privacy harms have not been established.°

Some big data developments, however, do not involve privacy concerns, such as news media monitoring databases that collect and make available previously-published and widely-disseminated news programming through a searchable database.° Even though news media monitoring databases do not raise privacy concerns, there is significant opposition to such databases, as the ongoing case of Fox News Network, LLC v. TVEyes, Inc. illustrates.° In TVEyes, several amici filed briefs in support of Fox’s copyright infringement claims against TVEyes, a news

---

° Institute’s Aspen Forum 1 (Aug. 19, 2013) (transcript available at: https://www.ftc.gov/sites/default/files/documents/public_statements/privacy-challenges-big-data-view-lifeguard%E2%80%99s-chair/130819bigdataaspen.pdf) (arguing that it should be the responsibility of the businesses who collect consumer data to make their methods clear to consumers, not hide them).

5. Elana Zeide, Student Privacy Principles for the Age of Big Data: Moving Beyond FERPA and FIPPS, 8 DREXEL L. REV. 339, 374 (2016) (identifying involuntary disclosure, information misuse and management, and repurposing as the major concerns relating to the collection of student information); Annelyse Gelman, Report from the Student Privacy Frontlines: 2015 in Review, EFF.ORG (Jan. 3, 2016), https://www.eff.org/deeplinks/2015/12/report-student-privacy-frontlines-2015-review (finding that ”school-issued Google Chromebooks upload private student data to the cloud by default, including web history”).


7. Thomas M. Lenard & Paul H. Rubin, Big Data, Privacy and the Familiar Solutions, 11 J.L. EXON. & POLY 1, 3–4 (2015) (arguing there is ”no evidence at present that big data used for commercial and other non-surveillance purposes has caused privacy harms” and privacy regulation would ”impose barriers to the innovation expected from the big-data revolution”); Adam Thierer, Who’s Governing Privacy? Regulations and Protection in a Digital Era, 66 M. L. REV. 467, 485 (2014) (”[O]utside narrow exceptions such as health and financial privacy regulation, the case for regulatory controls becomes harder to justify since the costs will typically exceed the benefits. To the extent greater information controls are pursued, the burden of proof lies with advocates of precautionary principle-based regulation to demonstrate unambiguous harms are omnipresent and unavoidable absent prophylactic constraints.”); Maureen K. Ohlhausen, The Procrustean Problem With Prescriptive Regulation, 23 COMM.LAW CONSPECTUS 1, 9 (2014) (advocating for ex post regulation for online privacy intrusions after harm has been established rather than ex ante regulation, in part, to avoid stunting technological development).


9. Id.
media monitoring database, including CNN,10 the National Association of Broadcasters,11 the National Cable and Telecommunications Association,12 and the National Press Photographers Association.13 On the other hand, several organizations have filed briefs in support of TVEyes’ news media monitoring services, such as the Computer and Communications Industry Association,14 the American Library Association,15 the Electronic Frontier Foundation,16 and Brave New Films.17

This Essay does not seek to resolve all of the specific disputes in TVEyes. Instead, this Essay has two purposes. One purpose is to emphasize that courts should be hesitant to deny the fair use rights of news media database creators because of the public benefits that these databases provide by making transformative uses of news content.18 Courts should be even more hesitant to deny the fair use rights of news media database creators when allegations of substitution market harm by copyright holders is speculative. Indeed, big data advocates urge caution against regulation when confronted with unproven privacy harms because such regulation could unnecessarily hinder technological development and the information economy.19 Such caution should be exercised when courts are confronted with unproven economic harm to news media outlets that seek to stunt the development of news

18. See infra Part III.
19. Ohlhausen, supra note 7, at 6 (describing how certain types of regulation, when not implemented correctly, would not be able to keep up with the fast pace of technology and would, among other things, hinder innovation).
media monitoring databases through the use of copyright infringement claims. The other purpose of this Essay is to provide preliminary thoughts on the role that the First Amendment should play when courts analyze the competing interests of copyright holders on the one hand, and creators and users of news media monitoring databases on the other. Thus, after providing background on the TVEyes case, two legal considerations are explored.

The first legal consideration is a core focus of the parties and amici in TVEyes, a fair use analysis.\(^\text{20}\) The second legal consideration, however, has not been addressed in the case by the parties or by the amici and it should receive attention when the legal permissibility of news media monitoring databases is considered.\(^\text{21}\) Specifically, the United States Supreme Court has recognized a “well-established,” albeit unclearly defined, qualified First Amendment right to receive information and ideas.\(^\text{22}\) Because of the constitutional role that the press plays in our self-governing, democratic society,\(^\text{23}\) the First Amendment right to receive information should not be ignored when analyzing the legal permissibility of news media monitoring database services.

Both the fair use and First Amendment right to receive information analyses indicate that courts should be reticent to uphold copyright infringement claims brought against creators of news media monitoring databases because of the important and transformative role these databases can play in our big data era, especially when any economic harm is speculative. Rather than allow copyright law to interfere with the development of news media monitoring databases generally, the appropriate remedy is either a direct copyright infringement claim against an end user of a news media monitoring database based on specific uses, or a secondary liability claim against the creator of a news media monitoring database based on allegations about the scope of specific functions of the database.\(^\text{24}\)

The creation of news media monitoring databases, the core functions of these databases, including the ability to view clips and indefinitely save clips, and the right of the general public to access these databases, are transformative uses of copyrighted news broadcasts that are protected fair uses. Moreover, news media monitoring databases serve vital First Amendment interests by allowing critical assessment of the Fourth Estate by researchers, media critics, other news outlets, and the general public. In other words, news media databases help hold the press

\(^{20}\) See infra Parts II, III.
\(^{21}\) See infra Part IV.
\(^{22}\) Stanley v. Georgia, 394 U.S. 557, 564 (1969) ("It is now well established that the Constitution protects the right to receive information and ideas" and that "[t]his right to receive information and ideas, regardless of their social worth is fundamental to our free society.").
\(^{23}\) David A. Anderson, The Origins of the Press Clause, 30 UCLA L. Rev. 455, 490–91 (1983) (describing the adversarial relationship between the press and democracy as a check on the government, one which a government may try to suppress, and as such freedom of press is a necessary element of self-government).
\(^{24}\) See infra Part III.
Joseph A. Tomain

accountable for their structural role in American constitutional democracy by using big data for publicly beneficial purposes.\textsuperscript{25}

II. \textit{FOX NEWS V. TVEYES BACKGROUND}

TVEyes is a for-profit, subscription-based, news media monitoring service that records over 1,400 television and radio stations 24 hours per day, seven days per week.\textsuperscript{26} It creates a keyword-searchable text index of the words spoken on these stations using speech-to-text technology and closed captioning.\textsuperscript{27} Fox does not challenge the creation of the keyword-searchable text index,\textsuperscript{28} but brought copyright infringement claims based on other aspects of TVEyes' services, including copying Fox's works to create the audiovisual database.\textsuperscript{29}

Subscribers pay a $500 monthly fee for access to the database.\textsuperscript{30} Subscribers include journalists, political campaigns, the White House, the United States Army, financial firms, and other political actors and businesses.\textsuperscript{31} Interestingly, TVEyes does not allow members of the general public to subscribe.\textsuperscript{32}

A variety of services are available to subscribers. A subscriber can: (1) search the database using keyword searches and view clips resulting from the search; (2) archive clips on TVEyes' servers; (3) download clips to its own computer; (4) share links to clips via email; and (5) search the database using a date/time search.\textsuperscript{33} TVEyes maintains the audiovisual recordings for only 32 days, but a subscriber can maintain indefinite access to specific clips by either saving clips in the archive on TVEyes' servers or downloading clips to their own computer.\textsuperscript{34} Clips are limited to no more than 10 minutes and TVEyes blocks a subscriber's ability to view more than 25 sequential minutes from a single station.\textsuperscript{35} Fox sued TVEyes for copyright

\textsuperscript{25} Keiyana Fordham, Note, \textit{Can Newspapers Be Saved? How Copyright Law Can Save Newspapers From the Challenges of New Media}, 20 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 939, 949–50 (2010) ("Where [media] ha[s] traditionally held government institutions accountable to the public, news aggregation websites have gained popularity and credibility by challenging mainstream media's dialogue with the public.").


\textsuperscript{27} Id.


\textsuperscript{29} Appellant Fox News Network, LLC at 3, Fox News Network, LLC v. TVEyes, Inc., No. 15-3885(L) & 15-3886(XAP) (2d Cir. June 16, 2016).

\textsuperscript{30} Fox News, 43 F. Supp. 3d at 385.

\textsuperscript{31} Id.

\textsuperscript{32} Id. at 385.

\textsuperscript{33} Id. at 384–85.

\textsuperscript{34} Id. at 385. Subscribers can also set up alerts for keywords and engage in other various analytics. Id. at 384–85.

\textsuperscript{35} Id. at 385.
Infringement, as well as state law claims. In multiple opinions, the district court analyzed each of these features for fair use, finding some features protected under the fair use doctrine and others not.

The court found the following features to be fair use: (1) copying the programs to create the audiovisual database; (2) allowing subscribers to search the database via keyword searches and to view or listen to the audiovisual clips; and (3) allowing subscribers to archive specific clips on TVEyes' servers indefinitely. On the other hand, the court found that the following features did not qualify as fair use: (1) allowing subscribers to download clips to their own computers; and (2) allowing subscribers to search the database using a date/time search. Finally, the court held that the share-by-email function qualified as fair use, so long as specific protections were instituted.

Both Fox and TVEyes appealed the district court decisions. At the time of writing, the case is pending before the United States Court of Appeals for the Second Circuit.

III. FAIR USE ANALYSIS

The main focus of the Fox News v. TVEyes litigation is whether the creation of the news media monitoring database and its related functions are acts of copyright infringement or are protected by the fair use doctrine. The Copyright and Patent Clause of the United States Constitution provides the source for copyright

---

36. The District Court held that Fox's two additional state misappropriation claims were preempted by the federal Copyright Act. Fox News, 43 F. Supp. 3d at 398–400. These misappropriation claims are not at issue on appeal.

37. See generally Fox News, 43 F. Supp. 3d 379.

38. Id. at 397.


41. Id.

42. Id. at 336–37.

43. Fox News Network, LLC v. TVEyes, Inc., 2015 WL 8148831, *1 (S.D.N.Y. Nov. 6, 2015). Those protections include, but are not limited to: limiting the number of email recipients outside of the subscribers' email domain who can receive a link to a clip; requiring the recipient to authenticate its email address with TVEyes; and providing clear and conspicuous notice that the clips are protected by copyright law. Id. at *1–2.


protection.” The constitutional purpose of copyright protection is to ensure that the public benefits from the creation of new works. To incentivize individuals to create new works, the Constitution provides creators of new works exclusive rights to those works for a limited time. Thus, the exclusive rights provided to creators of new works for a limited time are simply the private means to a public end.

Even during the period of the exclusive rights, however, fair use allows others to make certain uses of the work without authorization from, or compensation to, the author. Although Congress codified the concept of fair use in the Copyright Act of 1976, fair use as a judicial doctrine dates back to at least 1841. Indeed, the United States Supreme Court has recognized that since the “infancy of copyright protection, some opportunity for fair use of copyrighted materials has been thought necessary to fulfill copyright’s very purpose, ‘[t]o Promote the Progress of Science and useful Arts.’”

Analysis of the Copyright Act’s fair use provision shows that the creation of news media monitoring databases through copying audiovisual works for inclusion in such databases easily satisfies the statutory fair use factors. The amicus brief filed by the Computer and Communications Industry Association requests that the Second Circuit expressly articulate a bright line rule applicable to all search databases that it has already implicitly recognized: “the copying necessary to create

47. U.S. Const. art. I, § 8, cl. 8. “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.” Id.

48. See Eldred v. Ashcroft, 537 U.S. 187, 247 (2003) (Breyer, J., dissenting) (quoting H.R. Rep. No. 100-609 at 17 (1988)) (“Under the U.S. Constitution, the primary objective of copyright law is not to reward the author, but rather to secure for the public the benefits derived from the authors’ labors. By giving authors an incentive to create, the public benefits in two ways: when the original expression is created and . . . when the limited term . . . expires and the creation is added to the public domain.”).


51. See e.g., Robert L. Kelly, INTELLECTUAL PROPERTY § 11:9, Westlaw (database updated Dec. 2015). Libraries, archives, news reporting, teaching, and research benefit from limited exceptions to copyright infringement laws and are favored examples of fair use of copyrighted materials. Id.; Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 576 (1994) (holding that a rap group’s unauthorized use of another’s song within their own music is fair use); Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 608 (2d Cir. 2006).

52. Campbell, 510 U.S. at 576 (citing Folsom v. Marsh, 9 F. Cas. 342, 348 (CCD Mass. 1841)).

53. Id. at 575.

54. 17 U.S.C. § 107 (2012). The four factors are:

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

(2) the nature of the copyrighted work;

(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

(4) the effect of the use upon the potential market for or value of the copyrighted work.

Id.
BIG DATA AND THE FOURTH ESTATE

and operate a search database is a fair use as a matter of law.\textsuperscript{55} This request, however, goes further than what is argued here.\textsuperscript{56} The amicus brief filed by several intellectual property law scholars focuses specifically on TVEyes’ database, stating that “TVEyes goes above and beyond what fair use requires” in imposing limitations on its subscribers use of the database.\textsuperscript{57}

In relevant part, the preamble of the fair use statute states that “the fair use of a copyrighted work, including such use by reproduction in copies . . . for purposes such as criticism, comment, news reporting . . . or research, is not an infringement of copyright.”\textsuperscript{58} The creation of an audiovisual news media monitoring database falls squarely within the protections provided by the preamble of the fair use statute because it provides opportunities for criticism, comment, news reporting, and research on what is covered by these stations, what is not covered, and how it’s covered.\textsuperscript{59} Not only does the preamble of the fair use statute support the creation of news media monitoring databases, the four statutory factors also support the creation of such databases when the “results are weighed together.”\textsuperscript{60}

The first factor considers “the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes.”\textsuperscript{61} In analyzing this factor, the TVEyes district court focused on the transformative purpose of the news media monitoring database, just as the Second

\textsuperscript{55} Brief Amicus Curiae of the Computer & Communications Industry Association in Support of Defendant-Appellant-Cross-Appellee TVEyes at 3, Fox News Network, LLC v. TVEyes, Inc., No. 15-3885(L) & 15-3886 (XAP) (2d Cir. Mar. 23, 2016). The brief draws a distinction between copying for “invisible uses” (e.g. copying entire works to create a searchable database) and copying for “visible uses” (e.g. what is displayed to an end user in response to a search of a database). Id. at 6. “[W]hereas invisible uses inherently pass fair use muster, visible uses require a separate analysis to ensure that they satisfy Section 107’s requirements.” Id. at 17.

\textsuperscript{56} This Essay does not address whether copying entire works for the creation of any type of database is a fair use as a matter of law. Rather, this Essay focuses solely on the creation of news media monitoring databases. See supra Part I.


\textsuperscript{58} 17 U.S.C. § 107 (2012).


\textsuperscript{60} Campbell v. Acuff-Rose, 510 U.S. 569, 578 (1994) ("Nor may the four statutory factors be treated in isolation, one from another. All are to be explored, and the results weighed together, in light of the purposes of copyright.").

Circuit did in Google Books and HathiTrust. A use ‘can be transformative in function or purpose without altering or actually adding to the original work.’ Indeed, “the creation of a full-text searchable database is a quintessentially transformative use.” That Google Books and HathiTrust involve textual works and TVEyes involves audiovisual works is not a material distinction when considering the creation of searchable databases for purposes of fair use, especially at a time when video search is a “nascent and rapidly evolving area.” To ensure that the ability to comment, criticize, and research the news media is functional, searchable databases of news media video clips are as necessary, if not more so, than searchable databases of text. Audiovisual databases of news media, made possible by big data, are necessary because of the “vast and diffuse media environment” that exists today and because the “actual images and sounds depicted on television are as important as the news information itself – the tone of voice, arch of an eyebrow, or upturn of a lip can color the entire story, powerfully modifying the content.”

To be sure, an end user of a news media monitoring database may engage in uses of the database that rise to the level of direct copyright infringement; or, a database creator may be found secondarily liable for copyright infringement by, for example, allowing its database to serve as a substitute for the original broadcasts. But, these violations should be addressed in the context of direct copyright infringement claims against specific users of the database; or, secondary liability claims against database creators based on specific actions of the database creator, not through a broad, direct copyright infringement claim based on the creation of the audiovisual database itself. Such a broad use of copyright law threatens the opportunity that big

---

64. HathiTrust, 755 F.3d at 97. Other circuit courts have also upheld the creation of a searchable database that includes entire works under a fair use analysis. E.g., A.V. ex rel. Vanderhye v. iParadigms LLC, 562 F.3d 630 (4th Cir. 2009) (copying entire written works for a database to detect plagiarism); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007) (addressing a database of thumbnail images).
68. Fox News, 43 F. Supp. 3d at 392.
69. See generally Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 919 (2005) (holding that “one who distributes a device with the object of promoting its use to infringe copyright . . . is liable for the resulting acts of infringement by third parties”).
Big Data and the Fourth Estate

data provides for the development of news media monitoring databases. The transformative purposes of news media monitoring databases, including the ability to hold the news outlets accountable for their coverage or lack thereof on matters of public interest, outweigh general direct copyright infringement claims based merely on the creation of the database itself. As discussed below in the analysis of the fourth fair use factor, copyright holders must generally establish substitution market harm to defeat the fair use rights of another. The mere creation of a news media monitoring database does not establish substitution market harm as a matter of law.

Moreover, the for-profit nature of a news media monitoring service does not shift the analysis of this factor in favor of the copyright holder. The Supreme Court has made clear that the for-profit nature of an entity using another’s work is not a dispositive factor, but merely one consideration to be “weighed along with other[s] in fair use decisions.” The more transformative the use, the less significant the commercial nature of the user becomes. The Google Books court applied this Supreme Court guidance in finding that the first factor favored Google’s creation of its book database, even though Google is a for-profit entity. The same analysis should be persuasive in finding that the first factor weighs in favor of the creation of news media monitoring databases, regardless of whether the creator is a for-profit actor. Recent commentary also supports the view that the commercial nature of a database creator seeking protection under the fair use doctrine is not a particularly relevant factor, and certainly not a dispositive one.

---

71. Id. at 10.
72. See infra Part III.
73. Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 100 (2d. Cir. 2014) (“[T]he full-text search function does not serve as a substitute for the books that are being searched.”) (citing Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 591 (1994)); Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 614 (2d Cir. 2006)).
74. See A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 639 (4th Cir. 2009) (noting that “the fact that the disputed use of copyrighted material is commercial is not determinative in and of itself”).
76. Id. at 579.
77. Authors Guild v. Google, Inc., 804 F.3d 202, 217 (2d Cir. 2015), cert. denied, 136 S. Ct. 1658 (2016); Id. at 219 (rejecting the argument that “commercial motivation should outweigh a convincing transformative purpose.”).
The second factor – “the nature of the copyrighted work” – either weighs in favor of a finding of fair use in the creation of news media monitoring databases or does not favor either side to the dispute and therefore is of minor significance in this fair use analysis. Ideas are expressly excluded from copyright protection under the Copyright Act. Facts are also clearly excluded from copyright protection. Although news programs have copyrightable elements, non-fictional works are not at the heart of copyright protection. Because the main purpose of a news media monitoring database is to provide a research tool for commentary and criticism on news media programs and outlets, the nature of the underlying works ought to weigh in favor of a finding of fair use. Contrary to this perspective, the Second Circuit did not view the second factor as weighing in favor of the database creator in Google Books. Similarly, the TVEyes district court found that this factor did “not weigh for or against a finding a fair use.” In any case, “the second factor has rarely played a significant role in the determination of a fair use dispute.”

The third fair use factor assesses “the amount and substantiality of the portion used in relation to the copyrighted work as a whole.” This factor cannot weigh against a finding of fair use once the transformative purpose of the database is recognized under an analysis of the first fair use factor because the creation of the news media monitoring database requires copying entire broadcasts to fulfill the

82. Google Books, 804 F.3d at 220 (“While the copyright does not protect facts or ideas set forth in a work, it does protect that author’s manner of expressing those facts and ideas.”).
83. Stewart v. Abend, 495 U.S. 207, 237 (1990) ("In general, fair use is more likely to be found in factual works than in fictional works."); see also Swatch Grp. Mgmt. Servs. Ltd. v. Bloomberg L.P., 756 F.3d 73, 89 (2d Cir. 2014) ("It is well established that the "scope of fair use is greater with respect to factual works than non-factual works." (quoting New Era Publ’ns Int’l v. Carol Publ’g Grp., 904 F.2d 152, 157 (2d Cir. 1990)).
84. H.R. REP NO. 94-1476, at 5679 (1976) ("The fair use of a copyrighted work…for purposes such as criticism, comment, news reporting…or research, is not an infringement of copyright.").
85. Google Books, 804 F.3d at 220.
Big Data and the Fourth Estate

The transformative purpose of the database. The Second Circuit in Google Books and HathiTrust, as well as the district court in TVEyes, recognized the necessity of copying entire works to fulfill the purpose of the databases. Thus, like the second fair use factor, the third fair use factor is not as significant as the first or fourth factors in determining whether a news media monitoring database is fair use.

The Supreme Court has described the fourth fair use factor – “the effect of the use upon the potential market for or value of the copyrighted work” – as “undoubtedly the single most important element of fair use.” If another’s use of a copyright holder’s work serves as a substitute for the original work, then this factor weighs strongly against a finding of fair use because it would usurp the market for the original work and thus interfere with the incentive for individuals to create new works. Courts must consider not only the market effect of the specific defendant’s actions, but also the potential harm to the copyright holder if others engaged in similar activities as the defendant. But, not all economic market harm is relevant in analyzing the fourth fair use factor, only economic harm that results from a new use serving as a substitute for the original work. If, for example, a parody or a critical review creates market harm, that economic effect is irrelevant to an analysis of the fourth fair use factor because such harm is not a result of substitution for the original, but rather a general decline in interest for the original work. Finally, courts must balance the “benefit the public will derive if the use is permitted and the personal gain the copyright owner will receive if the use is denied.”

90. See Fox News, 43 F. Supp. 3d at 394–95.
91. Analyzing whether it is fair use to copy entire broadcasts for the purpose of creating a database is not the end of the analysis under the third fair use factor. In Google Books, the Second Circuit also considered whether the amount of material made available to database users as search results qualified as a fair use under the third factor. Google Books, 804 F.3d at 222–23. A similar issue arises in TVEyes: is the amount of material that a TVEyes’ subscriber can view in response to search results within the scope of fair use? This essay does not seek to resolve this specific issue. As in Google Books, however, this is a legitimate issue that must be considered when conducting a fair use analysis of a database comprised of entire works. Fox News Network, LLC v. TVEyes, Inc., 124 F.Supp.3d 325, 328 (S.D.N.Y. 2015).
95. Authors Guild Inc. v. HathiTrust, 755 F.3d 87, 95–96 (2d Cir. 2014).
98. Bill Graham Archives v. Dorling Kindersely Ltd., 448 F.3d 605, 613 (2d Cir. 2006) (quoting MCA, Inc. v. Wilson, 677 F.2d 180, 183 (2d Cir. 1981)).
Joseph A. Tomain

Properly designed news media monitoring databases do not serve as substitutes for the original works. Instead, these databases make a transformative use of the vast sea of news media broadcasts by creating a research tool for comment on and criticism of the news media, among other purposes. Big data provides the opportunity to engage in research, comment, and criticism of the news media in ways that would not be feasible without such technology. Additionally, when a copyright holder seeks to prevent the creator of a news media monitoring database from including its broadcasts in the database, it must rely on more than mere speculation to establish substitution market harm.

The TVEyes district court rejected Fox’s claims of market harm in its analysis of the fourth fair use factor because they were speculative. Not only did the TVEyes district court reject Fox’s claims of substitution market harm as speculative, it stated that any ‘‘cognizable market harm’ that can occur is likely to be outweighed by the public benefit arising from TVEyes’ services.”

In short, the TVEyes district court did exactly what big data proponents often advocate when faced with claims of alleged privacy harms caused by big data: it rejected speculative claims of harm, especially when uses of

99. See Fox News Network, LLC v. TVEyes, Inc., 43 F. Supp. 3d 379, 393 (S.D.N.Y. 2014) (quoting HathiTrust, 755 F.3d at 96) (”TVEyes’ search engine together with its display of result clips is transformative, and ‘serves a new and different function from the original work and is not a substitute for it.’”).
100. Id. at 395 (“[T]ransformative uses . . . ‘by definition do not serve as substitutes for the original work.’”)
101. Id. (quoting Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 99 (2d Cir. 2014)). Even if the use of another’s work is for a transformative purpose, however, substitution market harm is possible and courts must engage in an analysis of this factor. Authors Guild v. Google, Inc., 804 F.3d 202, 223–24 (2d Cir. 2015).
102. Brief of Media Critics as Amici Curiae in Support of Defendant-Appellant-Cross-Appellee at 24, Fox News Network, LLC v. TVEyes, Inc., Nos. 15-3885(L) & 15-3886(XAP) (2d Cir. Mar. 30, 2016) (“Absent the mass digitization facilitated by TVEyes’ database, there is a no feasible way for media critics to capture and present a comprehensive view of all content being broadcast to the public.”).
103. See Fox News, 43 F. Supp. 3d at 393.
104. Id. at 396–97. The court described the public benefits of TVEyes’ news media monitoring database as follows:

TVEyes subscribers use this service to comment on and criticize broadcast news channels. Government bodies use it to monitor the accuracy of facts reported by the media so they can make timely corrections when necessary. Political campaigns use it to monitor political advertising and appearances of candidates in election years. Financial firms use it to track and archive public statements made by their employees for regulatory compliance. The White House uses TVEyes to evaluate news stories and give feedback to the press corps. The United States Army uses TVEyes to track media coverage of military operations in remote locations, to ensure national security and the safety of American troops. Journalists use TVEyes to research, report on, compare, and criticize broadcast news coverage. Elected officials use TVEyes to confirm the accuracy of information reported on the news and seek timely corrections of misinformation. Clearly, TVEyes provides substantial benefit to the public.

Id. at 397.
Big Data and the Fourth Estate

big data provide substantial public benefits.\footnote{Id. at 397; cf. Brief for Amici Curiae Google Inc. and Microsoft Corp. at 18–19, Fox News Network, LLC v. TVEyes, Inc., No. 15-3885(L) & 15-3886 (XAP) (2d Cir. Mar. 23, 2016); Brief for the Electronic Frontier Foundation & Public Knowledge as Amici Curiae Supporting Appellant-Cross-Appellee TVEyes, Inc. at 22–24, Fox News Network, LLC v. TVEyes, Inc., No. 15-3885(L) & 15-3886 (XAP) (2d Cir. Mar. 23, 2016).} If speculative assertions of privacy harm by individuals are insufficient to support legal claims that risk obstructing the development of big data or disrupting the information economy, speculative assertions of economic harm by media conglomerates seeking to obstruct the development of news media databases through intellectual property claims should fare no better.

While the TVEyes district court found the creation of the audiovisual database itself, the ability to view clips, and the ability for a subscriber to archive clips indefinitely on TVEyes’ servers to be protected fair uses, it found other features not protected under the fair use doctrine.\footnote{Fox News, 43 F. Supp. 3d at 397–98.} The Second Circuit ought to reverse the district court and find at least two of these other features to be protected by the fair use doctrine. First, the ability to search news media monitoring databases by date and time is no less a fair use than the ability to search such databases via keyword searches. Among other reasons, date and time searches help discover the absence of coverage at a specific time and allow for a comparative analysis of news coverage across stations at a specific time.\footnote{Brief of Amici Curiae Professors of Intellectual Property Law in Support of Appellant/Cross-Appellee at 16, Fox News Network, LLC v. TVEyes, Inc., Nos. 15-3885(L) & 15-3886 (XAP) (2d Cir. Mar. 31, 2016); Brief for Amici Curiae Google Inc. and Microsoft Corp. at 27–28, Fox News Network, LLC v. TVEyes, Inc., Nos. 15-3885(L) & 15-3886(XAP) (2d Cir. Mar. 23, 2016).} Second, just as end users are able to indefinitely archive and access clips on TVEyes’ servers, they ought to be able to download those clips to their own computers. Being connected to the internet should not be a requirement of fair use. The mere ability to download clips indefinitely is not copyright infringement per se because fair uses can be made of downloaded clips, such as offline comparative research of multiple clips. Instead of focusing on whether the clips are stored on the database provider’s servers or its subscribers’ computers, copyright infringement claims must assess specific uses of the clips.

There are at least two other considerations in the context of a fair use analysis of news media monitoring databases that are not at issue in TVEyes, but that should also qualify as fair uses. First, a news media monitoring service that maintains recordings longer than 32 days without requiring user action should also be a fair use. Indeed, the TVEyes district court stated that “[c]ontent does not suddenly become unfit for fair use on the 33rd day after its creation.”\footnote{Fox News Network v. TVEyes, Inc., 124 F. Supp. 3d 325, 334 (S.D.N.Y. 2015).}

Second, a news media monitoring service made available to the general public should also be protected fair use. The TVEyes district court makes clear the
significant public interest provided by news media monitoring databases. Once it has been established that these databases provide significant benefits to the general public, it would be a questionable legal standard, at best, to draw a bright line rule that denies the general public access to such databases. The proper approach to balancing copyright owner’s rights and the fair use rights of others is either a direct copyright infringement claim against a user based on specific uses of the database or a secondary liability claim against the database creator based on specific allegations beyond the mere creation of the database or whom it allows access to the database. While some individuals or institutions may be able to provide social value through their use of news media databases based on their subject matter expertise, members of the general public may also be able to provide social value based on their own life experiences, or they may simply be exercising their qualified First Amendment right to receive information and ideas.

IV. FIRST AMENDMENT RIGHT TO RECEIVE INFORMATION

The press plays a structural role in our constitutional framework. A major function of the press is to serve as a watchdog holding the government and private actors accountable to the people. Thus, news media monitoring database services, made possible by big data, provide an individual the “crucial capability [to serve] as a ‘watchdog’ on the ‘watchdogs.’” Not only is there a policy reason for the necessity of allowing the people to serve as a watchdog on the watchdogs, there is, at least arguably, a constitutional basis for precluding copyright holders from using copyright law to prohibit the creation and use of news media monitoring databases.

In 1943, the United States Supreme Court first recognized that the First Amendment “necessarily protects the right to receive” information. By 1969, the Court stated that “[i]t is now well established that the Constitution protects the right to receive information and ideas.” Although this qualified First Amendment right to receive information and ideas has been “well established” since at least

1969, it “remains a relatively unexplored aspect of freedom of speech.”115 This constitutional right has arisen in a variety of contexts before the Court.116 But, the Court has not always found that this qualified right allows the requested speech or speech-related activity and it has not always been the main focus of the case.117 There are at least two Supreme Court cases, however, that support finding a First Amendment right to receive information through the use of news media monitoring databases: Red Lion Broadcasting v. FCC118 and Turner Broadcasting System, Inc. v. FCC.119

In Red Lion, the Court upheld the constitutionality of the fairness doctrine, which required broadcast radio and television stations to present each side of controversial public issues.120 Broadcasters challenging the fairness doctrine claimed that it violated their First Amendment rights by requiring them to provide coverage (i.e. engage in speech) that they might not choose to provide but-for the government’s mandate through the fairness doctrine.121 While the Court acknowledged that the broadcasters’ First Amendment rights were affected by the doctrine, it emphasized that it “is the right of the viewers and listeners, not the right of broadcasters, which is paramount.”122 Particularly relevant to the argument in this Essay, the Court also stated: “It is the right of the public to receive suitable access to social, political, esthetic, moral, and other ideas and experiences which is crucial here. That right may not constitutionally be abridged either by Congress or by the FCC.”123 Also relevant to this Essay, the Court stated that “[t]he freedom of the

117. See supra note 116.
120. 395 U.S. at 375. Although the fairness doctrine has been abolished by Congress, Red Lion’s holding that the doctrine was constitutionally permissible has not been overruled. Rosel H. Hyde, FCC Action Repealing the Fairness Doctrine: A Revolution in Broadcast Regulation, 38 SYRACUSE L. REV. 1175, 1175 (1987).
121. Red Lion, 395 U.S. at 386.
123. Red Lion, 395 U.S. at 390 (emphasis added).
press from governmental interference under the First Amendment does not sanction repression of that freedom by private interests.\textsuperscript{124}

In *Turner Broadcasting*, the Court upheld the constitutionality of “must-carry” legislation that required cable operators to include local broadcast stations in their services.\textsuperscript{125} Joining the majority on this point, Justice O’Connor pointed to the Court’s 1943 decision where it first recognized a First Amendment right to receive information.\textsuperscript{126} Additionally, and consistent with *Red Lion*, the Court emphasized that the “First Amendment’s command that government not impede the freedom of speech does not disable the government from taking steps to ensure that private interests not restrict, through physical control of a critical pathway of communication, the free flow of information and ideas.”\textsuperscript{127}

Both *Red Lion* and *Turner Broadcasting* demonstrate that there are times when the Court will recognize that citizens’ First Amendment rights to receive information and ideas outweigh the rights of content providers.\textsuperscript{128} The creation and use of news media monitoring databases is one of those times where the First Amendment rights of citizens to receive information and ideas outweigh the rights of content providers. Big data has revolutionized the ability of citizens and others to hold the Fourth Estate accountable for its performance as a watchdog.\textsuperscript{129} This newly-developed technological ability cannot be fulfilled without the right to receive the information made possible by the creation of news media monitoring databases.

The First Amendment did not sanction the ability of private actors to repress viewpoints on controversial public issues in *Red Lion*\textsuperscript{130} and it did not prevent the government from requiring that private actors allow access to their cable systems in *Turner Broadcasting*.\textsuperscript{131} In both cases, the First Amendment right of viewers and listeners to receive information and ideas outweighed the competing interests of the content providers.\textsuperscript{132} Similarly, the First Amendment rights of those that create and use news media monitoring databases outweigh the rights asserted by content providers seeking to prevent their content from being included in such databases.

\textsuperscript{124.} Id. at 392 (quoting Associated Press v. United States, 326 U.S. 1, 20 (1945)).
\textsuperscript{126.} Id. at 683 (O’Connor, J., concurring in part and dissenting in part) (citing Martin v. City of Struthers, Ohio, 319 U.S. 141, 149 (1943)) (noting that the government could not cut off solicitors’ access to homes where the owners were willing to accept them).
\textsuperscript{127.} Id. at 657 (citing Associated Press, 326 U.S. at 20).
\textsuperscript{128.} See id.; Red Lion, 309 U.S. at 390.
\textsuperscript{132.} *Turner*, 512 U.S. at 657; *Red Lion*, 395 U.S. at 390.
because of the unique opportunity that these databases provide for allowing the public and others to watch the watchdogs.

Of course, there is one significant difference between Red Lion and Turner Broadcasting on the one hand, and TVEyes on the other. The former cases involved balancing competing First Amendment rights of both sides of the dispute; whereas, the latter involves the scope of copyright protection afforded to those seeking to prevent inclusion of their content in a news media monitoring database. Indeed, neither party in TVEyes nor the amici raise First Amendment claims or defenses in the case. There is a practical reason likely explaining why First Amendment issues have not been raised by the parties and amici: Supreme Court precedent.

The United States Supreme Court has limited lower courts’ ability to engage in a First Amendment analysis in copyright cases by stating that free speech concerns are sufficiently addressed by “built-in free speech safeguards” within copyright law, such as fair use. Although the Court has stated that copyrights are not categorically immune from First Amendment challenges, it also stated that First Amendment scrutiny is unnecessary unless Congress has “altered the traditional contours of copyright protection.” On a surface level, this Supreme Court precedent is problematic from a pragmatic perspective of litigants that would like lower courts to analyze the First Amendment implications in copyright infringement cases. This precedent, however, does not provide a conclusive resolution of the interaction between the First Amendment and copyright law.

When confronted with the use of emerging technology in the big data era, especially in the context of news media monitoring databases which raise important First Amendment values of free speech and the role of the press, existing Supreme Court precedent is not directly on point. Indeed, the “traditional contours” of copyright law might be of little guidance when analyzing the intersection of the First Amendment and copyright law in this emerging digital context. Moreover, several scholars have criticized the Court’s seeming refusal to allow serious engagement with First Amendment concerns in copyright infringement cases. Finally, the common law process and the ability of the Court to clarify the applicability of prior precedent in new contexts, make consideration of First Amendment concerns necessary.

133. 395 U.S. at 386; 512 U.S. at 626–27; Fox News, 124 F. Supp. 3d 325.
Amendment concerns worthwhile when considering how technology has provided new opportunities to improve the free flow of information in our world and the ability for the public to assess the press as its agent through the use of big data.

V. CONCLUSION

The recently approved European Union’s General Data Protection Regulation provides that the “processing of personal data should be designed to serve mankind.”\(^{138}\) News media monitoring databases serve mankind because they allow individuals and others to critically examine how the press performs its role in holding the government and private actors accountable. Moreover, news media monitoring databases do not raise the privacy concerns involved in the processing of personal data because they are comprised of previously-published and widely-disseminated information. Thus, opposition to the creation and use of news media monitoring databases must be cautiously analyzed because these databases serve mankind. Indeed, U.S. courts also recognize the public benefits flowing from the creation of databases through opportunities arising in the big data era.

Google Books and HathiTrust show that the creation of searchable databases that copy entire works are quintessential transformative purposes that serve vital public interests.\(^{139}\) That news media monitoring databases involve copying audiovisual works and allow end users to view clips is not a material distinction, so long as they do not serve as market substitutes for the original works. In TVEyes, the Second Circuit ought to be skeptical of the speculative claims of economic harm that Fox asserts in an effort to overcome the fair use rights of TVEyes, especially in light of the significant public benefits that emerging news media monitoring databases offer. Just as big data proponents argue that speculative claims of privacy harm should not prevent the development of new uses of big data, speculative claims of economic harm should not prevent the development of news media databases.

More broadly, it is time for litigants, Congress, and policymakers to reconsider how First Amendment rights and copyright law intersect generally, and specifically in the context of news media monitoring databases. This analysis should include consideration of the qualified First Amendment right to receive information because this “well established” right remains underexplored, yet increasingly important in our big data era where technology provides new opportunities to analyze vast amounts of news content in ways that were not feasible in the past. While it may require the Supreme Court to revise its current approach to balancing


\(^{139}\) Authors Guild v. Google, Inc., 804 F.3d 202, 207–08 (2d Cir. 2015), cert. denied, 136 S. Ct. 1658 (2016); Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 97 (2d Cir. 2014).
speech and copyright interests, “it may be a long, a long time coming, but I [hope] a change gonna come.”

140. SAM COOKE, A CHANGE IS GONNA COME (RCA Victor 1964).