

# “HOW TWEET IT IS!”: HAVE TWITTER ARCHIVES BEEN LEFT IN THE DARK?

Seemantani Sharma<sup>†</sup>

## Abstract

*Social media is an increasingly prevalent method of communication. The information disseminated through these platforms is by nature ephemeral and at risk of loss. This has led institutions to build social media collections for posterity. The value of preserving social media for research purposes is increasingly important, yet significant legal issues must be addressed to make such collections viable.*

*While there is ample scholarly discourse on legal issues in web archiving, the same is not true for the newer sub-field of social media archiving. In this vein, this Article takes Twitter as a case study to analyze the potential legal issues that libraries and archives might encounter while developing a Twitter archive. Three issues were found to be most pertinent: (i) Copyright (ii) Privacy and (iii) Right of Publicity. While copyright is strictly a legal issue, privacy and right of publicity also have ethical paradigms to it calling for all stakeholders in the libraries and archives community to brainstorm.*

---

<sup>†</sup> Legal and Intellectual Property Services Officer of the Asia-Pacific Broadcasting Union, Kuala Lumpur (Malaysia); LL.M. (Intellectual Property), Class of 2015, The George Washington University Law School, Washington D.C. (U.S.); B.A. LL.B. (Hons.), Class of 2013, Amity Law School, Guru Gobind Singh Indraprastha University, New Delhi (India). The research for this Article was carried out in my capacity as the legal and policy researcher for the National Historical Publications & Records Commission (NHPRC) grant funded project Building Social Media Collections: Tools & Policy Recommendations for Collection Development & Management awarded to the George Washington University Libraries. My role on the project lasted for almost a year starting from April 2015 until May 2016. Preliminary findings of this research were presented at the 44th Research Conference on Communications, Information and Internet Policy organized by the Antonin Scalia Law School George Mason University, Arlington on September 30, 2016 and the Web Archives 2015: Capture, Curate, Analyze organized by the University of Michigan on November 12, 2015. I am grateful to the University of Michigan for providing me with a travel grant to attend the Web Archives 2015 in Ann Arbor, Michigan. The views reflected in this Article do not reflect the views of the George Washington University on the legal issues related to building social media collections and neither does it constitute binding legal advice. For specific legal advice or questions, please consult a licensed attorney in your jurisdiction. I am deeply grateful to Dan Chudnov, Vakil Smallen, Rachel Tent, Laura Wrubel, Justin Littman, Daniel Kerchner, Christie Peterson and the entire Social Feed Manager team for their support throughout the course of this project. All errors and omissions are purely mine. The author can be contacted at ssharma2@law.gwu.edu.

## TABLE OF CONTENTS

I.	Introduction.....	50
II.	Social Media Archives in United States.....	52
	A. LOC's Twitter Archive.....	52
	B. Social Media Archives by Universities .....	53
	C. Subscription-based Social Media Archives .....	54
III.	Legal Issues in Building a Twitter Archive .....	55
	A. Copyright.....	55
	1. Copyright Protection for Tweets .....	56
	a. Independent Origin .....	57
	b. Creativity .....	57
	c. Sufficiently Permanent or More Than Transitory Duration .....	59
	d. Fair Use .....	61
	B. The Purpose and Character of the Use .....	62
	1. Commercial or Non-profit Purposes.....	63
	2. Nature of the Use.....	63
	3. Transformative Use .....	63
	C. The Nature of the Copyrighted Work.....	64
	D. The Amount and Substantiality of the Portion Used .....	65
	E. The Effect of the Use Upon the Potential Market for or Value of the Copyrighted Work .....	66
	1. Privacy.....	68
	F. Right of Publicity .....	76
	1. If Celebrity's Image is Used for Advertising or Promoting the Twitter Archive.....	77
	2. If a Tweet Related to a Celebrity is Harvested. ....	77
IV.	Conclusion .....	78

## I. INTRODUCTION

Social media archiving is a newer sub-field of digital archiving which emerged in 2010 when the Library of Congress (LOC) exuberantly announced its ambitious plans of acquiring the entire Twitter archive for preservation purposes.<sup>1</sup> One of the avowed objectives of this project is to create a digital repository of publicly available tweets for posterity.<sup>2</sup> This announcement generated lot of publicity amongst the library and the archivist community, the tech community, and the academic and research community for its importance for digital heritage preservation.<sup>3</sup> The importance of this and other similar projects is monumental considering the ephemeral nature of social media, which

1. Matt Raymond, *How Tweet It Is!: Library Acquires Entire Twitter Archive*, LIBR. CONG. BLOG (Apr. 14, 2010), <https://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive>.

2. *Update on the Twitter Archive at the Library of Congress*, LIBR. CONG. BLOG (Dec. 2017) [hereinafter LIBR. CONG. BLOG], [https://blogs.loc.gov/loc/files/2017/12/2017dec\\_twitter\\_white-paper.pdf](https://blogs.loc.gov/loc/files/2017/12/2017dec_twitter_white-paper.pdf).

3. Kaitlin L. Costello & Jason Priem, *Archiving Scholars' Tweets*, SOC'Y AM. ARCHIVISTS (Apr. 2011), <http://www2.archivists.org/sites/all/files/KCFinal.pdf>.

is at risk of loss if not preserved.<sup>4</sup> However, before undertaking such a project, a risk assessment of legal issues must be undertaken by institutions involved in such endeavors for ensuring their viability.<sup>5</sup> While there is ample scholarly discourse on legal issues involved in web archiving, the same is not true for the newer sub-field of social media archiving. This Article is an attempt to address some of the legal issues that libraries and archives should keep in mind before creating a social media archive by taking Twitter as a case-study due to its widespread popularity as a valuable resource for researchers and academics.<sup>6</sup> It is because of this, that apart from the LOC, many other university libraries in the United States (U.S.) have launched counterparts of LOC's Twitter archive though of course their scale varies. Some of the general legal and ethical issues applicable for building a Twitter archive are also applicable to other social media platforms such as Facebook, Flickr, Instagram, Twitter, LinkedIn, YouTube and Sina Weibo.<sup>7</sup>

Part II of the Article gives a general introduction to the state of social media archives in the U.S. by categorizing it into three types: (i) LOC's Twitter archive,

---

4. Kalev Leetaru, *Why We Need To Archive The Web In Order To Preserve Twitter*, FORBES (July 18, 2017, 9:58 PM), <https://www.forbes.com/sites/kalevleetaru/2017/07/18/why-we-need-to-archive-the-web-in-order-to-preserve-twitter/#2dbe83e84b5e>.

5. In the context of web archiving, Brewster Kahle, founder of the Internet Archive has himself acknowledged the potential legal implications of web archiving. See Alyssa N. Knutson, *Proceed With Caution: How Digital Archives Have Been Left in the Dark*, 24 BERKELEY TECH. L.J. 437, 446 (2009) (arguing that there were legal and social issues in web archiving which most institutions cannot touch because it hits every privacy, copyright, and export controversy).

6. See Zeynep Tufekci, *Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls*, PROC. 8TH INT'L AAAI CONF. ON WEBLOGS & SOC. MEDIA 505, 506 (June 2014), <https://arxiv.org/ftp/arxiv/papers/1403/1403.7400.pdf> (stating that the preponderance of Twitter studies was mostly due to availability of data, tools and ease of analysis. Further, very large data sets, millions or billions of points were available from this source); see also Shea Bennett, *10 Must-Learn Lessons For Twitter Newbies*, ADWEEK (Aug. 6, 2012), <http://www.adweek.com/digital/10-lessons-twitter-newbies> (stating that Twitter was an amazing source of data); see also Klint Finley, *Twitter Opens Its Enormous Archives to Data-Hungry Academics*, WIRED (June 2, 2016, 4:41 PM), <https://www.wired.com/2014/02/twitter-promises-share-secrets-academia> (stating that academics and researchers saw huge value in the data collected by social media companies like Twitter and Facebook); see also *Social Media Harvesting Tools*, NCSU LIBR., <https://www.lib.ncsu.edu/social-media-archives-toolkit/collecting/social-media-harvesting-tools> (last visited Mar. 14, 2019) (stating that tweets were an emerging data source for researchers but due to the vast quantities of tweets, harvesting a usable data pool was difficult). Further, Twitter's popularity amongst the young, college-educated, and affluent Americans has prompted researchers in the social sciences to increasingly rely on it for studying important events. See Nick Ruest & Ian Milligan, *An Open-Source Strategy for Documenting Events: The Case Study of the 42<sup>nd</sup> Canadian Federal Election on Twitter*, 32 CODE{4}LIB. J. (2016), <http://journal.code4lib.org/articles/11358> (explaining that when historians study the 2015 Canadian federal election, Twitter will be a prime source). For an example of Twitter's role in the 2016 Presidential Election, see Bob Marcotte, *Twitter Researchers Offer Clues for Why Trump Won*, U. ROCHESTER NEWSCENTER (Feb. 20, 2017), <http://www.rochester.edu/newscenter/what-twitter-and-data-science-tell-us-about-the-2016-election-218762> ("Twitter, in particular, is a rich source of data because the millions of tweets posted by its members each day are easily accessible using an application programming interface."); see also Shannon Greenwood et al., *Social Media Update 2016*, PEW RES. CTR. (Nov. 11, 2016), <http://www.pewinternet.org/2016/11/11/social-media-update-2016> (stating that 21% of all U.S. adults used Twitter).

For the attitude of scholars on their tweets being archived, see Costello & Priem, *supra* note 2 (finding the opinions of 28 interviewees varied, with most people advocating that some tweets be saved for future use). For other such studies, see Chloé S. Georas, *Networked Memory Project: A Policy Thought Experiment for the Archiving of Social Networks by the Library of Congress of the United States*, 3(3) LAWS 469 (2014), <http://www.mdpi.com/2075-471X/3/3/469> (exploring the challenges of archiving social media posts).

7. Julie Tappendorf, *Social Media: Legal and Ethical Considerations*, ARCHIVESOCIAL (Sept. 3, 2015), <https://archivesocial.com/blog/social-media-legal-and-ethical-considerations>.

(ii) social media archives developed by universities such as the Social Feed Manager (SFM) developed by the George Washington (G.W.) University Libraries and the Social Media Archives Toolkit (SMAT) developed by the North Carolina State University (NCSU) Libraries, and (iii) subscription-based social media archives.

Part III of the Article analyzes the legal issues involved in building a Twitter archive. Three issues were found to be most pertinent: (i) Copyright, (ii) Privacy, and (iii) Right of publicity. This Section argues that while copyright and right of publicity are solely legal issues, privacy also has an ethical paradigm to it which transgresses into the sphere of archival theory and practice. Part IV of the Article concludes by arguing that while the legal framework for creating a Twitter archive is on firm grounds, the issue of ethics is far from clear. An absence of consensus on the acceptable ethical standards in creating web-based collections is required in order to develop workable solutions.

## II. SOCIAL MEDIA ARCHIVES IN UNITED STATES

### A. *LOC's Twitter Archive*

Social media archiving traces its history to a 2010 blog post by the LOC where it revealed its ambitious plans of creating a Twitter archive of every single public tweet since its inception in March 2006 in line with its mission of preserving the nation's digital heritage.<sup>8</sup> A white paper released by the LOC clearly articulated the importance of such an archive: "Archiving and preservation outlets such as Twitter will enable future researchers access to a fuller picture of today's cultural norms, dialogue, trends and events to inform scholarship, the legislative process, new works of authorship, education and other purposes."<sup>9</sup> The LOC's Twitter archive with its bona fide objective of preserving born digital records is unfortunately still in limbo due to lack of foresight on the part of ex-Librarian of Congress James Billington.<sup>10</sup> Apparently Billington could not visualize the magnitude of the project, especially the technological, financial and human resources required for the project.<sup>11</sup> Until August 2016, the LOC was still grappling with how to manage the archive which amounts to something like half a trillion tweets.<sup>12</sup> This undue delay had bred frustration amongst the researchers who were looking at LOC's Twitter archive as a low cost solution for seeking access to an otherwise exorbitantly priced Twitter's data sets.<sup>13</sup> However, in a recently released white paper, the LOC has

---

8. Raymond, *supra* note 1; see also Biz Stone, *Tweet Preservation*, TWITTER BLOG (Apr. 14, 2010), [https://blog.twitter.com/official/en\\_us/a/2010/tweet-preservation.html](https://blog.twitter.com/official/en_us/a/2010/tweet-preservation.html) (explaining how the Library of Congress is preserving tweets).

9. *Update on the Twitter Archive At the Library of Congress*, LIBR. CONG. (Jan. 2013), [https://www.loc.gov/static/managed-content/uploads/sites/6/2017/02/twitter\\_report\\_2013jan.pdf](https://www.loc.gov/static/managed-content/uploads/sites/6/2017/02/twitter_report_2013jan.pdf).

10. Nancy Scola, *Library of Congress' Twitter Archive is a Huge #Fail*, POLITICO (July 11, 2015, 5:09 PM), <https://www.politico.com/story/2015/07/library-of-congress-twitter-archive-119698.html>.

11. *Id.*

12. Andrew McGill, *Can Twitter Fit Inside the Library of Congress?*, THE ATLANTIC (Aug. 4, 2016), <https://www.theatlantic.com/technology/archive/2016/08/can-twitter-fit-inside-the-library-of-congress/494339>.

13. *Id.*

announced that instead of archiving all the tweets, it will only collect tweets from certain selected accounts based on the advice of its Collections Policy Statements and associated documents.<sup>14</sup> Most likely, it will collect and archive tweets which are thematic and event based such as the presidential elections and events of national importance.<sup>15</sup> This means that perhaps in near future researchers may finally be able to access the Twitter archive even though the white paper is silent about the timeline of the project.<sup>16</sup> All it says is that the Twitter archive “will remain embargoed until access issues could be resolved in a cost-effective and sustainable manner.”<sup>17</sup>

Apart from LOC’s Twitter archive, similar projects are also underway at both public and private university libraries such as the NCSU’s “Social Media Archives Toolkit” and Washington University’s “Documenting the Now” project (in collaboration with Maryland Institute for Technology in the Humanities).<sup>18</sup> These have been explored in-depth in the following sub-sections of the Article.

### B. Social Media Archives by Universities

The NCSU and the G.W. libraries are forerunners in harvesting social media to enable cultural heritage institutions to preserve social media at a low cost.<sup>19</sup> For accomplishing this, the NCSU libraries, based on an EZ Innovation Grant, have developed a tool known as Lentil for harvesting images from Instagram.<sup>20</sup> They have also developed a “Social Media Archives Toolkit” to enable other cultural organizations which are interested in collecting and curating social media to undertake similar social media archives program.<sup>21</sup> The NCSU libraries have done a good job at documenting their experiences in developing a social media archive including the legal and ethical implications of harvesting social media.<sup>22</sup>

The G.W. libraries with the financial support of the National Historical Publications and Records Commission (NHPRC) and the Institute of Museum and Library Services (IMLS) have developed an application known as the SFM for collecting datasets of social media records for aiding research and

---

14. LIBR. CONG. BLOG, *supra* note 2; *see also* Lizzie Plaugic, *The Library of Congress Will No Longer Archive Every Tweet*, THE VERGE (Dec. 26, 2017, 2:13 PM), <https://www.theverge.com/2017/12/26/16819748/library-of-congress-twitter-archive-project-stalled> (explaining how the Library of Congress will not archive tweets).

15. LIBR. CONG. BLOG, *supra* note 2.

16. *Id.*

17. *Id.*

18. *See Social Media Archives Toolkit*, NCSU LIBR. [hereinafter *Social Media Archives Toolkit*], <https://www.lib.ncsu.edu/social-media-archives-toolkit> (last visited Mar. 4, 2019) (describing the social media archive project); Julie Hale, *Documenting the Now: Archiving Social Media for Generations to Come*, WASH. U. ST. LOUIS (Jan. 20, 2016), <https://library.wustl.edu/8497-2>.

19. *Social Media Archives Toolkit*, *supra* note 18; *see also About the Project and Software*, SOC. FEED MANAGER G [hereinafter SOC. FEED MANAGER], <https://gwu-libraries.github.io/sfm-ui/about> (last visited Mar. 14, 2019) (talking about the social feed manager project and software).

20. NCSU LIBRARIES/LENTIL, <https://github.com/NCSU-Libraries/lentil> (last visited Mar. 14, 2019).

21. *Social Media Archives Toolkit*, *supra* note 18.

22. *Id.*

scholarship.<sup>23</sup> As of date, SFM only collects tweets, but in the future, it intends to harvest data from other social media sites such as Tumblr, Flickr, and Sina Weibo as well.<sup>24</sup> Just like NCSU, G.W. libraries have documented their experiences in developing the SFM quite well.<sup>25</sup> Their annotated bibliography on legal and ethical issues of social media collections is particularly useful for librarians, archivists and university lawyers who are involved in developing similar social media collections.<sup>26</sup>

Recently, the Andrew W. Mellon Foundation pledged a two year grant amounting to \$517,000 to fund a project known as “Documenting the Now: Supporting Scholarly Use and Preservation of Social Media Content” to be undertaken jointly by the Washington University in St. Louis, the Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland and the University of California, Riverside.<sup>27</sup> The aim of this project is to develop “a cloud-ready, open-source application that will be used for collecting tweets and their associated metadata and Web content.”<sup>28</sup> The project also seeks to produce a white paper on ethical, copyright and access issues related to collecting social media content.<sup>29</sup>

### C. *Subscription-based Social Media Archives*

Apart from the library created social media archives, there are also certain subscription-based social media archives such as Archive-It and ArchiveSocial which are popular for archiving social media.<sup>30</sup> Archive-It is an Internet Archive<sup>31</sup> affiliated tool which collects all online content including social media.<sup>32</sup> It can harvest social media content from Facebook, Twitter, and Instagram. Some of the institutions using it are the John Hopkins University, University of Wisconsin-Madison, Texas State Library and Archives

---

23. SOC. FEED MANAGER, *supra* note 19.

24. *Id.*

25. For the documentation, see SOC. FEED MANAGER, *supra* note 19 (explaining that Social Feed Manager is an open-source software).

26. See *Legal and Ethical Issues of Social Media Collecting: Annotated Bibliography*, SOC. FEED MANAGER (June 1, 2009) [hereinafter *Annotated Bibliography*], <https://gwu-libraries.github.io/sfm-ui/resources/annotated-bibliography> (providing an annotated bibliography).

27. *Documenting the Now: Archiving Social Media for Generations to Come*, WASH. U. ST. LOUIS (Jan. 20, 2016), <https://library.wustl.edu/8497-2>; see also DOCNOW [hereinafter *Documenting the Now*], <http://www.docnow.io> (last visited Mar. 4, 2019) (providing an example of efforts to collect and preserve digital content from social media chronicling historically significant events); see also Lisa Peet, *Documenting the Now Builds Social Media Archive*, LIBR. J. (May 2, 2016), <http://lj.libraryjournal.com/2016/05/academic-libraries/documenting-the-now-builds-social-media-archive> (describing the efforts undertaken by the Washington University and the Maryland Institute of Technology in the Humanities).

28. *Documenting the Now*, *supra* note 27.

29. *Id.*

30. *Social Media Harvesting Tools*, N.C. STATE U. RALEIGH [hereinafter *Social Media Harvesting Tools*], <https://www.lib.ncsu.edu/social-media-archives-toolkit/collecting/social-media-harvesting-tools> (last visited Mar. 14, 2019).

31. See *About*, INTERNET ARCHIVE, <https://archive.org/about> (last visited Mar. 14, 2019) (providing an example of a non-profit digital library of internet sites and other cultural artifacts in digital form, and noting the way that Internet Archive harvests web pages via a tool known as the Wayback Machine); see *The Way Back Machine*, INTERNET ARCHIVE, <https://archive.org/web> (last visited Mar. 4, 2019) (describing the way that Internet Archive uses the Way Back Machine).

32. *About Archive-It*, ARCHIVE-IT, <https://www.archive-it.org/learn-more> (last visited Mar. 4, 2019).

Commission (TRAIL), and the Oklahoma State Library's Oklahoma Digital Prairie collection.<sup>33</sup> On the other hand, ArchiveSocial collects and preserves content from Facebook, Instagram, Twitter, LinkedIn and YouTube and is commonly used by government agencies to comply with Freedom of Information Act (FOIA) requests.<sup>34</sup> It is also oriented towards the financial services sector for ensuring compliance with regulations enforced by the Securities and Exchange Commission (SEC) and the Financial Industry Regulatory Authority (FINRA).<sup>35</sup> Some of the institutions and governments using this service include the State Archives of North Carolina, the South Carolina State Library, the city of Austin, Texas and Snohomish County, Washington.<sup>36</sup>

All three projects are a stepping stone for preserving born digital records, which are at a risk of loss if not preserved.<sup>37</sup> However, their viability is not bereft of legal implications, some of which are discussed in the next Section in the context of Twitter.

### III. LEGAL ISSUES IN BUILDING A TWITTER ARCHIVE

#### A. Copyright

Copyright issues are the biggest obstacles facing digital libraries.<sup>38</sup> Due to this, any institution involved in designing a Twitter archive has to bypass the legal hurdle of copyright.<sup>39</sup> An institution can only be held liable for copyright infringement if there is copyright protection for tweets.<sup>40</sup> Even if tweets are subject to copyright protection, libraries and archives can make a respectable fair-use argument to ward off legal liability.<sup>41</sup> Both these issues have been discussed in the subsequent sections of the Article.

---

33. *Social Media Harvesting Tools*, *supra* note 30.

34. *Id.*; *see also Why Archive*, ARCHIVESOCIAL, <https://archivesocial.com/why-archive> (last visited Mar. 4, 2019) (highlighting the cooperative relationship that ArchiveSocial has with governments).

35. *See Annotated Bibliography*, *supra* note 26 (providing sources on legal issues regarding social media archiving and government agencies).

36. *See id.* (recording that these institutions make use of this service).

37. *See generally id.* (providing sources that note the risks of losing digital records without a concerted effort to preserve and archive them).

38. A government panel has found that copyright was the most significant barrier to preserving American cultural heritage in digital libraries. *See* Hannibal Travis, *Building Universal Digital Libraries: An Agenda for Copyright Reform*, 33 PEPP. L. REV. 761, 785 (2006) (noting the difficulties that copyright poses for the preservation efforts of libraries); *see also* Alyssa N. Knutson, *Proceed with Caution: How Digital Archives Have Been Left in the Dark*, 24 BERKLEY TECH. L.J. 437, 450 (2009) (stating that copyright dominated the digital archive discussion).

39. *See* Kyle-Beth Hilfer, *Tweet Tweet: Can I Copyright That?*, L. TECH. NEWS (Jan. 19, 2010), <http://www.webcitation.org/5ovHc7j9z> (pointing out some of the legal issues regarding copyrights and tweets).

40. *See id.* (noting the conditions under which a tweet might be copyrighted, and some defenses for those who nevertheless make use of the tweet).

41. *See id.* (describing the applicability of the fair-use defense, and the factors that a court would consider if presented with a fair-use defense as applied to the use of copyrighted tweets).

### 1. *Copyright Protection for Tweets*

In the U.S., the copyrightability of tweets has *per se* not been challenged before the courts.<sup>42</sup> However, in *Agence Fr. Presse v. Morel*, the copyrightability of photographs embedded as tweets was challenged before the United States District Court for the Southern District of New York.<sup>43</sup> This case arose in the aftermath of the 2010 Haitian earthquake when Daniel Morel, a native of Haiti clicked pictures of the devastation and uploaded thirteen of them to his new Twitter account.<sup>44</sup> Subsequently, Agence Fr. Presse (AFP) and Getty images misappropriated the uploaded images by distributing them without Morel's authorization.<sup>45</sup> AFP obtained the photographs from the Twitter feed of Lisandro Suero, who did not take the photos but was initially and mistakenly credited as the photographer.<sup>46</sup> Morel asserted that the distribution of the images infringed his copyright.<sup>47</sup> AFP sued Morel seeking a declaration that it had not infringed Morel's photographs as Twitter's Terms of Service permitted distribution of images without seeking consent.<sup>48</sup> The Court rejected AFP's argument that the Twitter's Terms of Service allowed it to distribute Morel's photos without permission including via "Getty images."<sup>49</sup> It held, "the Twitter TOS [Terms of Service] provides that users retain their rights to the content they post, with the exception of the license granted to Twitter and its partners."<sup>50</sup> Hence, the Court rebutted AFP's claim that Twitter intended to confer a license on it to sell Morel's photographs by a third-party stock photo agency, such as Getty images.<sup>51</sup> This case settles the issue that embedded media in a tweet, such as images, are subject to copyright protection.<sup>52</sup>

Initially, tweets were 140 characters in length but in November 2017, Twitter announced that tweets would be around 280 characters in length for users all across the world in order to facilitate ease of expression.<sup>53</sup> It is extremely unlikely that a 140-character tweet would be granted copyright protection as it will almost always lack the requisite level of originality to be

---

42. Michael L. Rustad, *Copyrights in Cyberspace: A Roundup of Recent Cases*, 12 J. HIGH TECH L. 106, 137 (2011).

43. *Agence Fr. Presse v. Morel*, 934 F. Supp. 2d 547 (S.D.N.Y. 2013).

44. *Id.* at 551.

45. *Id.* at 552–53.

46. *Id.* at 552.

47. *Id.* at 550.

48. *Agence Fr. Presse*, 934 F. Supp. 2d at 550.

49. *Id.* at 559–64.

50. *Id.* at 563.

51. *See id.* at 563 ("Moreover, even assuming that these statements grant *some* form of license to third parties ('other users') to, for example, re-tweet content posted on Twitter—a question not before the Court—they do not suggest an intent to grant a license covering the activities at issue here.")

52. *See also* Michael Keyes, *Trick or Tweet? Team Trump Gets Sued Over Skittles Twitter Pic*, DORSEY (Oct. 20, 2016), <https://thetmca.com/trick-or-tweet-team-trump-gets-sued-over-skittles-twitter-pic> (discussing the viability of a copyright claim against an image used on Twitter without permission).

53. *See* Casey Newton, *Twitter Is Rolling Out 280-character Tweets Around the World*, THE VERGE (Nov. 7, 2017, 3:56 PM), <https://www.theverge.com/2017/11/7/16616076/twitter-280-characters-global-rollout> (noting the change in the character allowances); *see also* Aliza Rosen, *Tweeting Made Easier*, TWITTER BLOG (Nov. 7, 2017), [https://blog.twitter.com/official/en\\_us/topics/product/2017/tweetingmadeeasier.html](https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html) (discussing the testing process for the change in tweet characters).



considered a copyrightable subject matter.<sup>54</sup> However, at 280 characters, there is a more reasonable probability that a tweet can be considered as a copyrightable subject matter provided it fulfils the originality criteria.<sup>55</sup> A work is said to be original if it is independently created and is creative.<sup>56</sup> Both these elements have been explored below:

a. Independent Origin

For a work to be original, it should be an independent creation of its author.<sup>57</sup> This means that the work should not have been the result of copying even though it resembles a prior work.<sup>58</sup> It does not mean that the work has to be novel in the sense that it should differ from other relevant works in some respect.<sup>59</sup> Therefore, as long as a tweet is not copied from another source, it can be deemed to be an independent work and can thus fulfill the independent origin criteria of copyrightability.<sup>60</sup>

b. Creativity

A work is said to be creative if it possesses an intellectual labor.<sup>61</sup> Only a modicum of creativity is required for a work to be copyrightable.<sup>62</sup> Therefore, for a tweet to be protected as a literary work, it must possess a minimum degree of creative spark irrespective of how crude or obvious it might be.<sup>63</sup> In *Applied Innovations, Inc. v. Regents of the Univ. of Minn.*, the United States Court of Appeals for the Eighth Circuit held that short, simple, declarative statements used in psychological tests may be protected by copyright as they were not

---

54. Under Section 102 of the U.S. Copyright law, a work is said to be original if it is independently created and has a minimum degree of creativity. See *Balt. Orioles, Inc v. Major League Baseball Players Ass'n*, 805 F.2d 663, 668 (7th Cir. 1986) (“The requirement of originality actually subsumes two separate conditions, i.e., the work must possess an independent origin and a minimal amount of creativity.”); see also *L. Baltin & Son, Inc. v. Snyder*, 536 F.2d 486, 490 (2d Cir. 1976) (“[I]t has been established that mass-produced commercial objects with a minimal element of artistic craftsmanship may satisfy the statutory requirement of such a work.”); *Withol v. Wells*, 231 F. 2d 550, 553 (7th Cir. 1956) (“A copyright protects an original work and is not dependent upon novelty.”). See generally Alan T. Dworkin, *Originality in the Law of Copyright*, 11 COPYRIGHT L. SYMP. 60 (1959) (laying down an in-depth analysis of the originality requirement).

55. See Terry L. Watt, *Legal Counsel: Can You Copyright a Tweet?*, NEWSOK (Aug. 19, 2018, 5:00 AM), <https://newsok.com/article/5605024/legal-counsel-can-you-copyright-a-tweet> (“In the case of the 280 character limit, it is more likely that such will be copyrightable. There are 277 characters in the three previous sentences. The longer tweet offers greater opportunity for creativity, so it is more likely to be protectable.”).

56. Dworkin, *supra* note 54, at 63.

57. *Id.* at 63.

58. See *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 346 (1991) (“[A]ssume that two poets, each ignorant of the other, compose identical poems. Neither work is novel, yet both are original and hence, copyrightable.”).

59. See *Baltimore Orioles*, 805 F.2d at 682 n.6 (explaining the difference between originality, creativity, and novelty).

60. Dworkin, *supra* note 56, at 63.

61. *Baltimore Orioles*, 805 F.2d at 682.

62. *Id.* at 669 n.7.

63. See *Feist Publ'ns, Inc.*, 499 U.S. at 345 (“The vast majority of works make the grade quite easily, as they possess some creative spark, ‘no matter how crude, humble or obvious’ it might be.”) (internal citation omitted).

fragmentary words, phrases, names, titles or slogans.<sup>64</sup> The U.S. Copyright Office (the Copyright Office) does not grant copyright protection to names, titles, short phrases and expressions.<sup>65</sup> Therefore, the initial 140-character tweet is likely to be considered as a short phrase or expression and probably outside the purview of copyright protection.<sup>66</sup> This is also because a significant proportion of the tweets are just updates about a person's current status, restatement of facts as to what they are doing and therefore, uncopyrightable.<sup>67</sup>

However, copyright protection for the newly introduced 280-character tweets cannot be ruled out.<sup>68</sup> This is specially so for creative tweets<sup>69</sup> such as short poems, short stories<sup>70</sup> and haikus<sup>71</sup> provided the fixation requirement is fulfilled.<sup>72</sup> A work is said to be fixed in a tangible medium of expression when

64. *Applied Innovations, Inc. v. Regents of the Univ. of Minn.*, 876 F.2d 626, 634–35 (8th Cir. 1989); *see Salinger v. Random House, Inc.*, 811 F.2d 90 (2d Cir. 1987) (stating though ordinary phrases “may be quoted without fear of infringement, a copier may not quote or paraphrase the sequence of creative expression that includes such [phrases]”).

65. U.S. COPYRIGHT OFF., COPYRIGHT BASICS 1, 2 (Sept. 2017), <https://www.copyright.gov/circs/circ01.pdf>; *see* Mary Minow, *Copyright Protection for Short Phrases—Rich Stim*, STAN. U. LIBR. (Sept. 9, 2003), [https://fairuse.stanford.edu/2003/09/09/copyright\\_protection\\_for\\_short](https://fairuse.stanford.edu/2003/09/09/copyright_protection_for_short) (explaining copyright protection for short phrases).

66. *See* John Lurie, *Twitter Copyright: Are Tweets Subject to Copyright Protection?*, LEGALMATCH (Mar. 22, 2016), <https://lawblog.legalmatch.com/2016/03/22/140-character-copyright-tweets-subject-copyright-protection> (acknowledging that “[s]hort phrases, such as tweets, can be protected under U.S. copyright law” but that “the higher standard for originality means that not every tweet will be protected.”).

67. *See* Mor Naaman et al., *Is it Really About Me? Message Content in Social Awareness Streams*, CSCW'10 PROCEEDINGS 2010 ACM CONF. ON COMPUTER SUPPORTED COOPERATIVE WORK 189, 192 (2010), <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.481.4600&rep=rep1&type=pdf> (showing that by dividing tweets into “meformers” and “informers,” the study found that a significant proportion of the tweets were “MeNow” messages which means that they were typically related to themselves or their thoughts); *see also* Lee Humphreys et al., *How Much is Too Much? Privacy Issues on Twitter*, PRIVACY ON TWITTER 1–2, <http://www.cs.utoronto.ca/~phillipa/papers/ica10.pdf> (finding that the majority of time, Twitterers write about themselves); Michael Beurskens, *Legal Questions of Twitter Research*, in TWITTER AND SOCIETY 125 (Katrin Weller et al. eds., 2013) (“Even though a layperson would probably consider a user to be the ‘author’ of a tweet, this does not necessarily imply that the tweet is actually protected by copyright law.”); *see also* Hilfer, *supra* note 39; Tim Rawson, *Framing it Another Way: Tweets, Copyright and the De Minimis Doctrine*, PILLSBURY INTERNET & SOC. MEDIA TEAM (Sept. 26, 2017), <https://www.internetandtechnologylaw.com/tweets-copyright-de-minimis-doctrine/> (“[M]any tweets consist of recited facts, which are not eligible for copyright protection.”).

68. Watt, *supra* note 55.

69. *Stern v. Does*, 978 F. Supp. 2d 1031, 1041 (C.D. Cal. 2011) (“Ultimately, the distinction between sentence and phrase is immaterial to the originality analysis. The focus must remain on the presence of creativity. While a shorter work, *ceteris paribus*, is less likely to possess the creative spark necessary to be accorded copyright protection, that will not always be the case.”). Further, the test is the degree of creativity involved in the text rather than its length. *See also* *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239, 251 (1903) (Harlan, J., dissenting) (arguing that copyright statute protection should not be withheld merely because a writing is short, as long as the other requirements for copyright protection are met); *Graphic Design Marketing Inc. v. Xtreme Enterprises Inc.*, 772 F. Supp. 2d 1029, 1032–33 (E.D. Wis. 2011) (holding that copyright protection for very short textual work depends upon its creativity. In this case, the word “STICKERS” was denied copyright protection on grounds that it was too common, short, and general to be copyrightable subject matter.); MELVILLE B. NIMMER & DAVID NIMMER, 1 NIMMER ON COPYRIGHT § 2.01(B) (1988) (“[T]he smaller the effort (e.g., two words) the greater must be the degree of creativity in order to claim copyright protection.”).

70. There are certain Twitter users such as Micro SF/F Fiction which compose fiction-tweet stories of 140 characters. *Micro SF/F Stories*, TWITTER (Mar. 28, 2018), <https://twitter.com/microsff?lang=en>; *see also* Ian Crouch, *The Great American Twitter Novel*, NEW YORKER (July 23, 2014), <http://www.newyorker.com/books/page-turner/great-american-twitter-novel> (discussing short stories written on Twitter).

71. *See Haiku*, KIDZONE, <http://www.kidzone.ws/poetry/haiku.htm> (last visited Mar. 5, 2017) (showing that Haiku is a traditional form of Japanese poetry consisting of three lines).

72. *Religious Tech. Ctr. v. Lerma*, 95-1107-A, 1996 WL 633131, at \*4 (stating *in dicta* that some short phrases and haikus may be copyrightable); *see also* Benjamin Beck & Konstantin von Werder, *Tweets Reported*

it is embodied in a copy or phonorecord, by or under the authority of the author, is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration.<sup>73</sup> Both these elements have been explored below.

Tweets cannot be viewed without an internet enabled device such as a desktop, laptop, mobile or tablet.<sup>74</sup> Therefore, the very act of posting a tweet leads to its fixation and copyright protection begins at the moment of fixation.<sup>75</sup>

### c. Sufficiently Permanent or More Than Transitory Duration

Every second, an average of around 6,000 tweets are tweeted on Twitter, which corresponds to over 350,000 tweets per minute, 500 million tweets per day and around 200 billion tweets per year.<sup>76</sup> This essentially means that a user's Twitter feed gets updated constantly.<sup>77</sup> Even with this continuous stream of tweets, they do not get deleted automatically nor do they overwrite the old or previously posted tweets.<sup>78</sup> Therefore, an originally posted tweet is more or less permanent unless it is deleted by the user.<sup>79</sup> Even otherwise, fixation does not imply permanence.<sup>80</sup> What is important is that the work can be identified in a medium "capable of identification" irrespective of whether it has a permanent endurance or not.<sup>81</sup> A tweet can easily be identified in a user's Twitter feed and lasts for more than a transitory duration.<sup>82</sup> Therefore, the fixation requirement for a tweet to become a copyrightable subject matter is fulfilled.<sup>83</sup>

In practice, courts do not judge the artistic merit of a work while assessing its copyrightability.<sup>84</sup> Therefore, practically, this discretion lies with the

*as Infringing Copyright Deleted by Twitter-Are Tweets Copyrightable?*, MAYER BROWN (Aug. 4, 2015), <https://www.allaboutipblog.com/2015/08/tweets-reported-as-infringing-copyright-deleted-by-twitter>; Adam S. Nelson, *Tweet Me Fairly: Finding Attribution Rights Through Fair Use in the Twittersphere*, 22 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 697, 718 (2012) (analyzing copyright protection for short phrases).

73. 17 U.S.C. § 101 (2018).

74. *But see* T.J. Ortenzi, *How to Use Twitter When You Lose Internet Access*, WASH. POST (Oct. 29, 2012), [https://www.washingtonpost.com/business/technology/how-to-use-twitter-when-you-lose-internet-access/2012/10/29/e6214f12-21b9-11e2-8448-81b1ce7d6978\\_story.html?utm\\_term=.6d9b3586497c](https://www.washingtonpost.com/business/technology/how-to-use-twitter-when-you-lose-internet-access/2012/10/29/e6214f12-21b9-11e2-8448-81b1ce7d6978_story.html?utm_term=.6d9b3586497c) (detailing that an SMS process might be used to connect to Twitter when offline).

75. Nelson, *supra* note 72, at 725.

76. *Twitter Usage Statistics*, INTERNET LIVE STATS, <http://www.internetlivestats.com/twitter-statistics> (last visited Mar. 5, 2019).

77. *See id.* (noting the number of new tweets per second).

78. *See* Yaara Lancet, *Did You Know Twitter Deletes Your Old Tweets? Here's How to Back Them Up*, MUO (Sept. 17, 2012), <https://www.makeuseof.com/tag/twitter-deletes-tweets-back-up> (noting that Twitter only allows access to a user's latest 3,200 tweets).

79. I say more or less permanent as I am not quite sure for how long Twitter stores an originally posted tweet on its servers. This should be distinguished from the shelf-life of a tweet which is not more than eighteen minutes. *See* Laura Sheptoski, *Are You Maximizing The Shelf Life of Your Social Media?*, WEIDERT GROUP (June 3, 2014), [https://www.weidert.com/whole\\_brain\\_marketing\\_blog/bid/206554/are-you-maximizing-the-shelf-life-of-your-social-media](https://www.weidert.com/whole_brain_marketing_blog/bid/206554/are-you-maximizing-the-shelf-life-of-your-social-media) (noting the lifetime of tweets).

80. Douglas Lichtman, *Copyright as a Rule of Evidence*, 52 DUKE L.J. 683, 717 (2003).

81. MELVILLE B. NIMMER & DAVID NIMMER, 1 NIMMER ON COPYRIGHT § 1.08(C)(2)(a) (1963).

82. *See* Sheptoski, *supra* note 79 (noting the average lifetime of a tweet).

83. *See* Evan Brown, *Fixed Perspective: The Evolving Contours of the Fixation Requirement in Copyright Law*, 10 WASH. J.L. TECH. & ARTS 17, 19 (2014) (noting the fixation requirement under the Copyright Act).

84. *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239, 251–52 (1903) ("It would be a dangerous undertaking for persons trained only to the law to constitute themselves final judges of the worth of pictorial illustrations, outside of the narrowest and most obvious limits . . . . Yet if they command the interest of any

Copyright Office.<sup>85</sup> In 2014, the Copyright Office refused to register Tweet #452—“Monkey bar fallacy: a bad person using something makes it bad. E.g., users of monkey bars include: children, TERRORISTS #tor”—on grounds that no copyright subsists in words and short phrases such as names, titles, and slogans even if they were distinctively arranged or printed.<sup>86</sup> The office action further stated that the tweet represented “less than the required minimum amount of original authorship on which to base a claim.”<sup>87</sup> At the face of it, perhaps this tweet has at least some degree of creativity. The arrangement of the words is distinct and plausibly has a hidden essence.<sup>88</sup> Therefore, whether a tweet is creative or not is subject to the discretion of the Copyright Office.<sup>89</sup>

From the above analysis, it is clear that the vast majority of tweets cannot be copyrighted on grounds of size, content and *scènes à faire* issues but protection for sufficiently creative tweets cannot be ruled out.<sup>90</sup> Even Twitter recognizes copyright protection for original and creative tweets.<sup>91</sup> In 2015, when Olga Lexell, a freelance writer, complained to Twitter about her tweets being republished by someone else without her permission, Twitter promptly deleted the infringed tweets by posting the message: “this tweet has been withheld in response to a report from the copyright holder.”<sup>92</sup> Lexell requested Twitter to intervene on grounds that she made her living writing jokes.<sup>93</sup> Perhaps Twitter honored her request after realizing that her livelihood was threatened

---

public, they have a commercial value—it would be bold to say that they have not an aesthetic and educational value—and the taste of any public is not to be treated with contempt.”)

85. See generally Robert K. Walker & Ben Depoorter, *Unavoidable Aesthetic Judgments in Copyright Law: A Community of Practice Standard*, 109 NW. U. L. REV. 343 (2015) (noting how artistic judgments are unavoidable in copyright law).

86. Gabriel J. Michael, *Can You Copyright a Tweet?*, TO PROMOTE THE PROGRESS? (July 30, 2014), <https://topromotetheprogress.wordpress.com/2014/07/30/can-you-copyright-a-tweet/>; Letter from U.S. Copyright Office on Tweet #452 to Gabriel Michael (July 23, 2014) [hereinafter Letter from U.S. Copyright Office], [https://topromotetheprogress.files.wordpress.com/2014/07/tweet\\_reg\\_denied.pdf](https://topromotetheprogress.files.wordpress.com/2014/07/tweet_reg_denied.pdf).

87. See Letter from U.S. Copyright Office, *supra* note 86.

88. I tried contacting the author of the tweet to understand its precise meaning but was without any luck.

89. See *supra* notes 84–85 and accompanying text (illustrating this discretion).

90. Consuelo Reinberg, *Are Tweets Copyright Protected?*, WIPO MAG. (July 2009), [http://www.wipo.int/wipo\\_magazine/en/2009/04/article\\_0005.html](http://www.wipo.int/wipo_magazine/en/2009/04/article_0005.html).

91. See Abby Minns, *US Copyright Infringement is No Joke on Twitter*, OSBORNE CLARK (Sept. 17, 2015), <http://marketinglaw.osborneclarke.com/media-and-ip/us-copyright-infringement-is-no-joke-on-twitter> (noting how Twitter removed questionable tweets).

92. See *id.* (discussing Twitter’s response to a Digital Millennium copyright Act take down notice by removing tweets containing a joke on the basis of copyright infringement); *Twitter Starts Deleting Tweets of Stolen Jokes Over Copyright Infringement*, ECON. TIMES (July 27, 2015, 9:47 PM) [hereinafter *Twitter Starts Deleting Tweets of Stolen Jokes*], <https://economictimes.indiatimes.com/tech/internet/twitter-starts-deleting-tweets-of-stolen-jokes-over-copyright-infringement/articleshow/48241410.cms?inttarget=no>. The removal of the infringing tweets was in sync with Twitter’s Terms of Service. See *Twitter Terms of Service*, TWITTER [hereinafter *Twitter Terms of Service*], <https://twitter.com/en/tos#usContent> (last visited Mar. 5, 2019) (discussing Clause 3 of the Terms of Service which states, “[w]e reserve the right to remove the Content that violates the User Agreement, including for example, copyright or trademark violations, impersonation, unlawful conduct, or harassment.”). The genesis of the debate over copyright protection for tweets dates back much before the Olga Lexell controversy. It owes its origin to a blogpost posted by Mark Cuban, the owner of Dallas Mavericks in 2009 where he questions whether a tweet could be copyrighted. See *Are Tweets Copyrighted?*, BLOG MAVERICK: THE MARK CUBAN WEBLOG (Mar. 29, 2009), <http://blogmaverick.com/2009/03/29/are-tweets-copyrighted> (questioning whether tweets are copyrightable).

93. *Twitter Starts Deleting Tweets of Stolen Jokes*, *supra* note 92.

due to this act of infringement and the fact that she was the legitimate copyright owner of her tweet.<sup>94</sup>

Since the vast majority of tweets are not copyrightable, harvesting them for the purposes for creating a Twitter archive will not entail legal liability.<sup>95</sup> However, if at all libraries and archives are sued for harvesting the few tweets in which copyright may subsist, they can make a respectable fair use argument, which has been explored below.<sup>96</sup> The same holds true for other not-for-profit institutions which are involved in building social media collections.<sup>97</sup>

#### d. Fair Use

Section 108 of the Copyright Act, 1976 allows libraries and archives to reproduce and distribute certain copyrighted works without permission on a limited basis for the purposes of preservation, replacement, and research.<sup>98</sup> However, this exception is severely restricted for the digital age as it was enacted during the analog era.<sup>99</sup> On April 29, 2015, the then Register of Copyrights, Maria A. Pallante, testified before the House Judiciary Committee and explained that the library exceptions were outdated to the point of being obsolete.<sup>100</sup> She also explained that the Copyright Office would be preparing a discussion draft for addressing structural and substantive changes to Section 108.<sup>101</sup> In line with her testimony, the Copyright Office commenced a review of Section 108 during the summer of 2016.<sup>102</sup> It held a series of nearly 40 in-person and telephonic meetings with various stakeholders such as librarians, archivists, museum professionals, content creators, scholars, and technology professionals.<sup>103</sup> The discussion draft does not propose an exception for web archiving at this stage because of the complexity of issues involved.<sup>104</sup> It says that a more detailed study was required particularly on issues related to evolution of technology, the notice-and-takedown process, meaning of publicly available works and

---

94. *Twitter Terms of Service*, *supra* note 92 (providing that Clause 3 of the Terms of Service states, “You retain your rights to any Content you submit, post or display on or through the Services. What’s yours is yours—you own your Content . . .”).

95. See Justin Littman, *Social Media Harvesting Techniques*, SOC. FEED MANAGER (Oct. 28, 2015), <https://gwu-libraries.github.io/sfm-ui/posts/social-media-harvesting-techniques> (noting some of the social media harvesting techniques).

96. See 17 U.S.C. § 107 (2018) (detailing the list of enumerated purposes which act as guidelines for fair-use arguments).

97. *Id.*

98. 17 U.S.C. § 108 (2018).

99. *Revising Section 108: Copyright Exceptions for Libraries and Archives*, COPYRIGHT.GOV, <https://www.copyright.gov/policy/section108> (last visited Mar. 5, 2019).

100. MARIA A. PALLANTE, THE REGISTER’S PERSPECTIVE ON COPYRIGHT REVIEW 14–15 (Apr. 29, 2015), [https://judiciary.house.gov/wp-content/uploads/2016/02/114-22\\_94408.pdf](https://judiciary.house.gov/wp-content/uploads/2016/02/114-22_94408.pdf).

101. See *id.* at 15 (stating that the discussion draft would “address museums, preservation exceptions and the importance of ‘web harvesting’ activities”).

102. U.S. COPYRIGHT OFF., SECTION 108 OF TITLE 17: A DISCUSSION DOCUMENT OF THE REGISTER OF COPYRIGHTS 12 (Sept. 2017), <https://www.copyright.gov/policy/section108/discussion-document.pdf>.

103. *Id.*

104. See *id.* at 42 (“After considering the broad range of issues that such statutory exception for web archiving would entail, the Copyright Office is not proposing an exception for the preservation and distribution of internet content at this time.”).

treatment of commercial works.<sup>105</sup> This statement is bit odd considering that even after a lapse of nine years since the first time a recommendation for an exception for libraries and archives permitting them to preserve publicly available online content was made, the Copyright Office has yet to study the complexities related to the issue.<sup>106</sup> The Section 108 Study Group had recommended way back in 2008 that “[a] new exception should be added to section 108 to permit libraries and archives to capture and reproduce publicly available online content for preservation purposes, and to make those copies accessible to users for purposes of private study, scholarship, or research.”<sup>107</sup> An opt out provision allowing rights holders to opt out of allowing libraries and archives (save with the exception of the LOC) to capture their publicly available online content was also made.<sup>108</sup> If this recommendation is implemented, it would enable libraries and archives to capture publicly available online content including social media content and to create a permanent copy of it for their collections.<sup>109</sup> Until this recommendation is enacted into law, libraries and archives have no option but to seek recourse to the fair use exception.<sup>110</sup> Section 107 of the Copyright Act permits use of a copyrighted work by way of reproduction or in copies or phonorecords or by any other means for purposes such as criticism, comment, news reporting, research, teaching, and scholarship.<sup>111</sup> The list of enumerated purposes act only as a guideline and not as a complete list of uses that are fair.<sup>112</sup> In determining whether the use of a copyrighted work is a fair-use or not, the following four factors are considered by the courts:<sup>113</sup>

### *B. The Purpose and Character of the Use*

The first factor of the fair use test is the purpose and character of the use of the copyrighted work.<sup>114</sup> To determine this, the courts look into three elements: (a) Whether the work is for commercial or non-profit purpose, (b) Whether the use is for criticism, comment, news reporting, teaching, scholarship or research, and (c) Whether the use is transformative or whether it puts the work to a new and substantially different use.<sup>115</sup>

---

105. *Id.*

106. U.S. COPYRIGHT OFF., NAT’L DIG. INFO. INFRASTRUCTURE & PRES. PROGRAM LIBR. CONG., THE SECTION 108 STUDY GROUP REPORT 80 (Mar. 2008) [hereinafter THE SECTION 108 STUDY GROUP], <http://www.section108.gov/docs/Sec108StudyGroupReport.pdf>.

107. *Id.* at viii.

108. *Id.*

109. *Id.* at 81.

110. *Id.* at 63–64.

111. 17 U.S.C. § 107 (2018).

112. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994).

113. *Id.*

114. *Id.*

115. *See id.* at 579 (discussing the transformative use element).

### 1. *Commercial or Non-profit Purposes*

Whether a use is for a commercial or non-profit purposes is only one of the factors that must be weighed in light of all other elements.<sup>116</sup> A Twitter archive designed by a university library or any other not-for-profit institution such as the LOC, which makes a dataset of tweets available to researchers either in raw form or otherwise, will fulfill this sub-factor of fair use until the time they do not charge a fee for making the collections available to patrons or unfairly profit from using the tweets.<sup>117</sup>

### 2. *Nature of the Use*

The second sub-factor is the nature of the use for which the copyrighted work is being used.<sup>118</sup> If the intended usage is for criticism, comment, news reporting, teaching, scholarship or research, it will likely be considered as fair use.<sup>119</sup> The intended usage enables courts to determine whether the defendant acted in good faith or bad faith.<sup>120</sup> A Twitter archive which aids scholarship, research and teaching will rank very high on this sub-factor since the Twitter archive is merely a tool to facilitate research.<sup>121</sup> A university library or not-for-profit institution has no commercial interest or for that matter any other mala fide interest in developing the Twitter archive.<sup>122</sup>

### 3. *Transformative Use*

Transformative use is a relatively new addition to the fair-use law.<sup>123</sup> It was raised for the first time in *Campbell v. Acuff-Rose Music*, where the U.S. Supreme Court held that a commercial parody could qualify as fair-use.<sup>124</sup> A work is said to be transformative when it either transforms the original work into a different character or purpose or alters the original work with new expression, meaning, or message.<sup>125</sup> Where the original work is used as “raw material” to create “new information, new aesthetics, new insights and understandings[,]” the secondary work is transformative.<sup>126</sup> Harvesting tweets via a software such

---

116. *Id.* at 584.

117. *Harper & Row, Publs. v. Nation Enters.*, 471 U.S. 539, 562 (1985); *see also* *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 932 (2d Cir. 1994) (holding that a private, for-profit corporate library could not rely on fair-use for systematically making unauthorized copies of copyrighted materials published in scientific journals for its employees).

118. *Campbell*, 510 U.S. at 577.

119. *Id.* at 576.

120. *Harper*, 471 U.S. at 562–63 (stating “Fair use presupposes ‘good faith’ and ‘fair dealing’” and “distinguishes between ‘a true scholar and a chiseler who infringes a work for personal profit’”) (internal citations omitted).

121. *Campbell*, 510 U.S. at 577.

122. Jacey Norris, *Art or Artifice: The Second Circuit’s Misapplication of the Fair Use Factors in Cariou v. Princein Light of Kienitz v. Sconnie Nation*, 25 DEPAUL J. ART, TECH. & INTELL. PROP. L. 429, 436 (2015) (“At first glance, one may categorize any unauthorized taking as a ‘bad faith’ taking, but there are different motivations for borrowing from a copyrighted work.”).

123. *Campbell*, 510 U.S. at 594.

124. *Id.*

125. 17 U.S.C. § 107 (2018); *Harper*, 471 U.S. at 562.

126. Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1111 (1990).

as the SFM and subsequently indexing it for creating a collection is sufficiently transformative.<sup>127</sup> This is because the tweets are akin to a “raw material” which are transformed or put to new use in the form of a research tool for enabling scholars, researchers, and students with their research or for the preservation of ephemeral social media content. Without the aid of these tools, researchers cannot readily find tweets tweeted under a particular hashtag. For example, a Ph.D. student studying the 2014 Ferguson unrest can get only get limited access to a small portion of the tweets under the hashtag “#blacklivesmatter” via Twitter’s Application Programming Interface (API).<sup>128</sup> The Twitter API would work well if the researcher does not intend to undertake a rigorous analysis of the data.<sup>129</sup> However, for undertaking a rigorous statistical analysis of the data, the researcher would have to either manually retrieve the tweets which would be very time consuming or else gain access to Twitter’s firehose which is quite expensive.<sup>130</sup> In *Kelly v. Arriba Soft Corp.*, the United States Court of Appeals for the Ninth Circuit held that the use of thumbnails in search engines results was transformative as it improved access to information.<sup>131</sup> Therefore, in light of this precedent since a Twitter archive also improves access to information, i.e., research material in the form of tweets, it is likely to weigh heavily on the transformative criterion. Further, social media archiving is akin to web archiving, which commentators have found to be sufficiently transformative.<sup>132</sup>

### C. *The Nature of the Copyrighted Work*

The second factor of the fair use test is “the nature of the copyrighted work.”<sup>133</sup> It “calls for recognition that some works are closer to the core of intended copyright protection than others, with the consequence that fair use is more difficult to establish when the former works are copied.”<sup>134</sup> Two types of distinctions have emerged for evaluating this factor: (1) whether the work is expressive or creative rather than predominantly factual; and (2) whether the work is published or unpublished.<sup>135</sup> Creative works receive a higher degree of protection than factual or informational works as law envisages them to be

---

127. See *Version 1.11.0*, GITHUB, <https://github.com/gwu-libraries/sfm-ui/releases> (last visited Mar. 15, 2019) (providing the source code for the Social Feed Manager).

128. See *Twitter Firehose vs. Twitter API: What’s the Difference and Why Should You Care?*, BRIGHTPLANET (June 25, 2013) [hereinafter *Twitter Firehose vs. Twitter API*], <https://brightplanet.com/2013/06/twitter-firehose-vs-twitter-api-whats-the-difference-and-why-should-you-care> (discussing Twitter API).

129. *Id.*; see *Evaluate Twitter Data To Inform Business Decisions*, TWITTER, <https://developer.twitter.com/en/use-cases/analyze> (last visited Mar. 15, 2019) (explaining Twitter’s API).

130. *Twitter Firehose vs. Twitter API*, *supra* note 128.

131. 336 F.3d 811, 822 (9th Cir. 2002).

132. See David M. Ray, *The Copyright Implications of Web Archiving and Caching*, 14 SYRACUSE SCI. & TECH. L. REP. 1, 25 (2006) (stating that “the archiving of web pages is transformative as it put the content it captures to a new use, which is the archival and display of websites for the purposes of historic preservation.”).

133. *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 586 (1994).

134. *Id.*

135. *Blanch v. Koons*, 467 F.3d 244, 256 (2d Cir. 2006).



public property.<sup>136</sup> Most tweets fall low on the creativity threshold as they are mostly plain mundane facts.<sup>137</sup> Therefore, the balance is likely to tip in favor of a Twitter archive under this factor. However, in the rare case of creative tweets, libraries and archives will have a high burden of proof under this factor.<sup>138</sup>

#### D. *The Amount and Substantiality of the Portion Used*

Under the third factor, the courts look into the amount and substantiality taken from the original copyrighted work.<sup>139</sup> This means how much does the secondary work borrow from the original work.<sup>140</sup> Courts do not always give this factor much weight as a work that borrows excessively from another may anyway run into trouble under the first and fourth factors.<sup>141</sup> A secondary work which consists predominantly of unoriginal elements will not be sufficiently transformative as there will be miniscule addition or alteration to the secondary work.<sup>142</sup> Such a work also poses an economic threat of serving as a perfect substitute of the original work.<sup>143</sup>

With respect to a Twitter archive, there are two ways in which this sub-factor comes into play.<sup>144</sup> First is when all the tweets from a handle are harvested and subsequently distributed to a researcher as a compilation.<sup>145</sup> In this case, this sub-factor may disfavor libraries and archives if majority of the tweets could be protected as literary works.<sup>146</sup> However, this is a rare possibility as most of the tweets are nothing but plain mundane facts such as what is someone doing etc.<sup>147</sup> The only exception to this could be tweets from handles such as Micro SF/F Fiction which compose fiction stories of 140 characters or tweets from those handles which develop a unique and consistent theme or voice.<sup>148</sup> Secondly, if a Twitter archive harvests tweets only under a particular hashtag,

---

136. *Id.*; see Norris, *supra* note 122, at 437 (arguing that “works of fiction or great creativity generally receive higher protection than factual or information works because the law conceptualizes facts as public property but recognizes that individuals have a property interest in their original expressions.”).

137. See Mor Naaman et al., *supra* note 67, at 192 (stating that by dividing tweets into “meformers” and “informers,” the study found that a significant proportion of the tweets were “MeNow” messages which means that they were typically related to themselves or their thoughts); Akshay Java et al., *Why We Twitter: Understanding Microblogging Usage and Communities*, SPRINGER 1, 2 (Aug. 12, 2007), [https://ebiquity.umbc.edu/\\_file\\_directory\\_/papers/369.pdf](https://ebiquity.umbc.edu/_file_directory_/papers/369.pdf) (stating that most tweets are “daily chatter, conversations, sharing information and reporting news.”).

138. See *supra* Part III (A)(i) (covering such aspects in this part of this Note).

139. *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 577 (1994).

140. *Id.* at 586–87.

141. Norris, *supra* note 122, at 438.

142. *Campbell*, 510 U.S. at 588.

143. See *Folsom v. Marsh*, No. 4901, 1841 U.S. App. LEXIS 468, at \*1–7 (holding that the fourth factor disfavors a finding of fair use if the market is impaired because the quoted material supersedes the use of the original).

144. *Campbell*, 510 U.S. at 588.

145. See Pascal Jürgens & Andreas Jungherr, *A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis*, CREATIVE COMMONS 1 (2016), <http://ssrn.com/abstract=2710146> (discussing the collection of Twitter data).

146. See generally *id.* (commenting on Twitter archive libraries).

147. Tim Soulo, *What Do People Tweet About & The Surprising Truth About What Drives Them*, BLOGGER JET (2016), <https://bloggerjet.com/what-do-people-tweet-about>.

148. @MicroSFF, TWITTER, <https://twitter.com/microsff?lang=en> (last visited Mar. 15, 2019); see Nelson, *supra* note 72, at 723 (discussing the characters in tweets).

then liability under this factor will anyway not arise as the archive will harvest the entire 140 character tweet rather than few characters of the tweet, which would be meaningless.<sup>149</sup> Therefore, on public policy grounds, this sub-factor tilts in favor of libraries and archives considering that courts are guided by striking a balance between protecting the works of authors and promoting the public interest.<sup>150</sup>

*E. The Effect of the Use Upon the Potential Market for or Value of the Copyrighted Work*

The fourth factor has been declared by the Supreme Court as “the single most important element of fair-use.”<sup>151</sup> At a basic level, this factor delves into whether a copyright owner would lose revenue should a fair use defense succeed.<sup>152</sup> There are two areas of concern under this sub-factor.<sup>153</sup> One is the market’s reaction to the secondary use and whether the reaction interferes with the original copyright owner’s market.<sup>154</sup> A not-for-profit Twitter archive would have a negligible effect on the potential market value as majority of the tweets are anyway not copyrightable.<sup>155</sup> This means that a Twitter archive cannot interfere with the original copyright owner’s market as there is no copyright ownership in a majority of the tweets in the first place.<sup>156</sup> For tweets which can be protected as literary works, the effect on the potential market would also be negligible as a non-commercial Twitter archive would only collect tweets from a particular hashtag or handle for aiding researchers.<sup>157</sup> Ideally, archives and libraries will not charge their patrons for this service. However, if they do decide to charge a service fee from the patrons, the courts are unlikely to find fair-use under this factor as it would hurt the revenues of the copyright owner, i.e., if at all.<sup>158</sup>

In light of the settlement in *Internet Archive v. Shell*, the applicability of the fair-use provisions for not-for-profit digital archives has been left in the dark.<sup>159</sup> However, from the above analysis, it is clear that libraries and archives

---

149. Nelson, *supra* note 72, at 723.

150. See U.S. CONST. art. I, § 8, cl. 8 (stating that the purpose of copyright law is “[t]o 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.”).

151. *Campbell*, 510 U.S. at 574.

152. Kelly Jines-Story, *Does Rocky & Bullwinkle Hold the Key to Unlocking the Mystery of Fair Use in the Age of Internet Archiving*, 35 CAP. U. L. REV. 1023, 1055 (2007).

153. Leval, *supra* note 126, at 1124 (arguing that by definition every fair use involves some loss of royalty revenue because the secondary user has not paid royalties).

154. *Id.*

155. See *supra* Part III (A)(i) (explaining the negligible impact of such an archive).

156. *Sony Corp. of Am. v. Universal City Studios*, 464 U.S. 417, 450 (1984) (holding that without a “demonstrable effect upon the potential market” there was no effect in prohibiting the secondary use).

157. Andres, *Can you Copyright a Tweet*, TECHNOLLAMA (Jan. 17, 2015), <https://www.technollama.co.uk/can-you-copyright-a-tweet>.

158. THE SECTION 108 STUDY GROUP, *supra* note 106.

159. See *Internet Archive v. Shell*, 505 F. Supp. 2d 755, 760 (D. Colo. 2007) (showing that whether fair use provisions are applicable to not-for-profit digital archives was not decided by the court in *Shell*); see also Brewster, *Lawsuit Settled*, ARCHIVE.ORG (Apr. 25, 2007), <https://archive.org/post/119669/lawsuit-settled> (announcing the settlement of the *Shell* lawsuit); Knutson, *supra* note 38, at 449 (arguing that the settlement

can make a respectable fair-use argument for developing a not-for-profit Twitter archive solely for research purposes. For strengthening the fair use argument, libraries and archives should comply with the Association of Research Libraries' Code of Best Practices (the Code) on the creation of web-based collections.<sup>160</sup> As per the Code, captured material should be represented as it was captured, with appropriate information on the mode of harvesting along with the date.<sup>161</sup> This is not an onerous requirement and as a matter of fact, the SFM periodically documents the mode and the date of harvesting the tweets.<sup>162</sup> Further, it also calls upon libraries and archives to identify the legal proprietors of the site in question.<sup>163</sup> In the context of a Twitter archive, this should again not be problematic as the exported spreadsheet of the harvested tweets can plausibly spell out the name of the Twitter handle and the Twitter ID from which the tweet was harvested.<sup>164</sup>

Lastly, the Code advises libraries and archives to give an opportunity to copyright owners for registering their objections for making items from web-based collections available online and that they should respond to such objections promptly.<sup>165</sup> This requirement may be redundant as university libraries and archives would ideally not make the Twitter archive available online. If at all they decide to make it available online, I would assume that access to the online collection would require user authentication. As far as responding to objections is concerned, this would depend upon the magnitude of the project. In the case of LOC's Twitter archive, it is hard to say whether the LOC would be able to dedicate a full-time staff to respond to objections. In the case of comparatively small-scale projects such as G.W.'s SFM and NCSU's SMAT, I think responding to objections would be feasible as it may not require a full-time staff dedicated to the job. Apart from copyright, developing a Twitter archive also entails an assessment of privacy concerns of Twitter users.

---

between Internet Archive and Suzanne Shell in *Internet Archive v. Shell* had led to an unpredictable fair use regime for digital archives).

160. See ASS'N RES. LIBR., AM. U. WASH. C. L., CODE OF BEST PRACTICES IN FAIR USE FOR ACADEMIC AND RESEARCH LIBRARIES 26 (Jan. 2012) [hereinafter CODE OF BEST PRACTICES], <http://www.arl.org/storage/documents/publications/code-of-best-practices-fair-use.pdf> (providing Association of Research Libraries' (ARL) Code of Best Practices relating to creation of web-based collections).

161. *Id.* at 27.

162. See *Overview of Social Feed Manager*, SOC. FEED MANAGER [hereinafter *Overview of Social Feed Manager*], <https://gwu-libraries.github.io/sfm-ui/about/overview> (last visited Mar. 15, 2019) (providing the mode of harvesting tweets); GITHUB, <https://github.com/gwu-libraries/sfm-ui> (last visited Mar. 15, 2019) (detailing the code). Further, it even generates some statistics for the datasets captured, see *A Peek at 251,077,140 #election2016 Tweets*, SOC. FEED MANAGER, <https://gwu-libraries.github.io/sfm-ui/posts/2017-10-11-election> (last visited Mar. 11, 2019) (explaining how it even generates some statistics for the datasets captured).

163. *Overview of Social Feed Manager*, *supra* note 162.

164. See *id.* (exemplifying how the Social Feed Manager enables researchers to export the tweets to a spreadsheet which contains the Twitter ID, the screen name, the hashtag and the Twitter handle of the harvested tweet).

165. See CODE OF BEST PRACTICES, *supra* note 159 (showing the options available to copywrite owners in these situations).

### 1. *Privacy*

Deciphering privacy related issues in building a Twitter archive is important since, as of date, Twitter's Terms of Service does not explicitly permit developing a Twitter archive for research and scholarly purposes though it does mention that Twitter can use tweets posted by users.<sup>166</sup> Notable information privacy scholar Michael Zimmer has raised grave privacy concerns related to LOC's Twitter archive.<sup>167</sup> He raises a controversial but thought provoking question that even though the LOC's Twitter archive will only archive publicly available tweets, it makes accessibility and searchability of tweets much easier whenever it is processed by the LOC.<sup>168</sup> Similar sentiment has been expressed by other commentators as well.<sup>169</sup> Despite these concerns, from a legal standpoint there is a grim possibility of this issue cropping up unless libraries and archives harvest protected tweets or deleted tweets.<sup>170</sup> This is because as per precedent, there is no reasonable expectation of privacy in a publicly posted social media content.<sup>171</sup>

In *People v. Harris*, the Court of Appeals of New York held that there was no reasonable expectation of privacy in a publicly posted tweet, as posting a tweet was analogous to screaming out of a window, a situation in which there is no reasonable expectation of privacy.<sup>172</sup> This case arose in the aftermath of the Occupy Wall Street movement where Malcolm Harris was charged with disorderly conduct. The District Attorney's office sent a *subpoena duces tecum* to Twitter seeking Harris's account information and tweets since they were relevant for the on-going criminal investigation.<sup>173</sup> Harris moved to quash the

---

166. For Twitter's Privacy Policy, see *Twitter Privacy Policy*, TWITTER, <https://twitter.com/en/privacy> (last visited Mar. 15, 2019) ("In addition to information you share with us, we use your Tweets, content you've read, Liked, or Retweeted, and other information to determine what topics you're interested in, your age, the languages you speak, and other signals to show you more relevant content.").

167. Michael Zimmer, *Open Questions About Library of Congress Archiving Twitter Streams*, MICHAEL ZIMMER (Apr. 14, 2010), <http://www.michaelzimmer.org/2010/04/14/open-questions-about-library-of-congress-archiving-twitter-streams>.

168. *Id.*; see also Michael Zimmer, *How Your Private Tweets Might Be Included in the Library of Congress Public Archive*, MICHAEL ZIMMER (Apr. 14, 2010), <http://www.michaelzimmer.org/2010/04/14/how-your-private-tweets-might-be-included-in-the-library-of-congress-public-archive> (showing how private tweets can be included in the Library of Congress Archive); Jasmine McNealy, *The Privacy Implications of Digital Preservation: Social Media Archives and the Social Networks Theory of Privacy*, 3 ELON L. REV. 133, 142 (2012) (exploring the privacy consideration of social media archives). The deed of gift signed between Twitter and LOC expressly states that only publicly available tweets will be archived. See *Gift Agreement*, LIBR. CONG. BLOG, <http://blogs.loc.gov/loc/files/2010/04/LOC-Twitter.pdf> (last visited Apr. 7, 2019) (explaining how the deed of gift signed between Twitter and LOC expressly states that only publicly available tweets will be archived).

169. See Catherine Lovrics, *Copyright and Privacy Questions Around Your Public Tweets and the New Library of Congress Archive and Google Replay*, SLAW: CANADA'S ONLINE LEGAL MAG. (Apr. 27, 2010), <http://www.slaw.ca/2010/04/27/copyright-and-privacy-questions-around-your-public-tweets-and-the-new-library-of-congress-archive-and-google-replay> (exploring the privacy concerns surrounding the new Library of Congress Archive).

170. See *About Public and Protected Tweets*, TWITTER, <https://help.twitter.com/en/safety-and-security/public-and-protected-tweets#> (last visited Mar. 16, 2019) (explaining that protected tweets or private tweets are those which may only be visible to a Twitter user's followers).

171. See, e.g., *People v. Harris*, 949 N.Y.S.2d 590, 593 (N.Y. Crim. Ct. 2012) (holding that there is no reasonable expectation of privacy in a publicly posted tweet).

172. *Id.* at 595.

173. *Id.* at 592.

subpoena on the ground that it violated his privacy rights under the Fourth Amendment and the provisions of the Stored Communications Act.<sup>174</sup> Twitter subsequently stated that it would not comply with the subpoena until the motion to quash was ruled on.<sup>175</sup> However, the court held that Harris had no standing to quash the subpoena as there was no proprietary interest in a publicly posted tweet since it was now gifted to the world.<sup>176</sup> It further held that a tweet was “not the same as a private email, a private direct message, a private chat, or any of the other readily available ways to have a private conversation via the Internet which now exist[ed].”<sup>177</sup> Further, even when a user deleted his or her tweets, there were “search engines available such as ‘Untweetable,’ ‘Tweleted[,]’ and ‘Politwoops[,]’ [which held] users accountable for everything they had publicly tweeted and later deleted.”<sup>178</sup> Hence, there was no violation of Harris’s Fourth Amendment rights as he had “no reasonable expectation of privacy in the information he [had] intentionally broadcasted to the world.”<sup>179</sup>

In a similar case, the Supreme Court of New York held that when the Plaintiff created her Facebook and MySpace accounts, she had consented to her personal information being shared with others, irrespective of the privacy settings.<sup>180</sup> In *Romano v. Steelcase Inc.*, the Supreme Court of New York rejected the plaintiff’s objection for giving access to her Facebook and MySpace accounts, including all deleted pages and related information on privacy grounds.<sup>181</sup> The court held that sharing personal information with others was “the very nature and purpose of social networking sites.”<sup>182</sup>

Lastly, in *Moreno v. Hanford Sentinel, Inc.*, the Court of Appeals of California held that there was no expectation of privacy in a publicly posted ode to MySpace.<sup>183</sup> In this case, a young woman penned down “An Ode to Coalinga” to her MySpace page while she was visiting her hometown Coalinga, California during her college holidays.<sup>184</sup> The ode despised her hometown and its residents.<sup>185</sup> Six days later, she deleted the post.<sup>186</sup> However, during this six day period, the principal of the Coalinga High School discovered it and sent it to the editor of a local newspaper.<sup>187</sup> The ode was published by the editor with the author’s last name appended.<sup>188</sup> Due to this, the author and her family received death threats and a shot was fired at the family home, forcing the family to move out of Coalinga.<sup>189</sup> The girl and her family sued the publishers of the

---

174. *Id.*

175. *Id.*

176. *Harris*, 949 N.Y.S.2d at 595.

177. *Id.*

178. *Id.*

179. *Id.*

180. *Romano v. Steelcase Inc.*, 907 N.Y.S.2d 650, 656 (N.Y. Sup. Ct. 2010).

181. *Id.*

182. *Id.*

183. *Moreno v. Hanford Sentinel, Inc.*, 91 Cal. Rptr. 3d 858, 862 (Cal. Ct. App. 2009).

184. *Id.* at 861.

185. *Id.*

186. *Id.*

187. *Id.*

188. *Moreno*, 91 Cal. Rptr. 3d at 861.

189. *Id.*

local newspaper, alleging invasion of privacy and intentional infliction of emotional distress.<sup>190</sup> The court held that the principal did not invade the author's privacy by handing down the post to the editor and that the editor did not violate her rights by publishing her name in the newspaper.<sup>191</sup> This is because material posted to MySpace diminished the user's expectation of privacy.<sup>192</sup> The court rejected her argument that only a limited audience would have viewed her posting.<sup>193</sup> The court held:

Here, Cynthia publicized her opinions about Coalinga by posting the Ode on MySpace.com, a hugely popular Internet site. Cynthia's affirmative act made her article available to any person with a computer and thus opened it to the public eye. Under these circumstances, no reasonable person would have had an expectation of privacy regarding the published material.<sup>194</sup>

In line with this precedent, since Twitter users have no reasonable expectation of privacy in their publicly available tweets, archiving them for scholarly and research purposes does not raise any legal concerns from the standpoint of privacy.<sup>195</sup> However, the same does not hold true for deleted tweets, protected tweets, or public tweets, which are subsequently deleted, as there is a reasonable

---

190. *Id.*

191. *Id.* at 862–63.

192. *Id.*

193. *Moreno*, 91 Cal. Rptr. 3d at 863.

194. *Id.* at 862.

195. See David J. Pittenger, *Internet Research: An Opportunity to Revisit Classic Ethical Problems in Behavioral Research*, 13(1) ETHICS & BEHAV. 45, 49 (2003) (arguing that the Internet offers no veil of privacy because any disclosure made through any Internet forum is by nature public); see also Catherine C. Marshall & Frank M. Shipman, *Social Media Ownership: Using Twitter as a Window Onto Current Attitudes and Beliefs*, PROCEEDINGS OF CHI. 1081, 1085 (2011) (discussing a survey respondent stating that while his Facebook was private, his Flickr, Twitter, and blog were public, giving a general sense of publicness of tweets); Ian Convery & Diane Cox, *A Review of Research Ethics in Internet Based Research*, 6 (1) PRAC. RES. HIGHER EDUC. 50, 52 (arguing that there may be fewer obligations to protect privacy if the research focuses on publicly accessible archives or environments intended by their authors or members to be public); Joseph B. Walther, *Research Ethics in Internet-Enabled Research: Human Subjects Issues and Methodological Myopia*, 4 (3) ETHICS & INFO. TECH. 205, 207 (2002) (arguing that expecting privacy in Internet communications is misplaced as it is well understood that it is open to public scrutiny); LEANNE TOWNSEND & CLAIRE WALLACE, *SOCIAL MEDIA RESEARCH: A GUIDE TO ETHICS* 1, 12 (2016) (noting that the issue of hashtags implied that Twitter users were keen to contribute to a community or to a public debate and therefore expected even more people to see their data).

See *Twitter Terms of Service*, *supra* note 92 (stating that by submitting, posting or displaying the tweets, users give Twitter a worldwide, non-exclusive, royalty-free license to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute the tweets); see also McNealy, *supra* note 168, at 136 (stating that under traditional privacy law, one could argue that internet users have no expectation of privacy in the information which they disclose on a social networking site). For privacy expectations in online social media, see Steven D. Zansberg & Janna K. Fischer, *Privacy Expectations in Online Social Media: An Emerging Generational Divide?*, 28 (3) COMM. LAW (2011), [http://www.lskslaw.com/documents/evolvingprivacyexpectations\(00458267\).pdf](http://www.lskslaw.com/documents/evolvingprivacyexpectations(00458267).pdf); see also *Four Navy Seals v. Associated Press*, 413 F. Supp. 2d 1136, 1143 (S.D. Cal 2005) (holding that there was no expectation of privacy in images uploaded to a publicly accessible website); *State v. Birchfield*, 2007 WL 1437235, at \*8–10 (N.J. Super. Ct. App. Div. May 17, 2007) (holding that there was no reasonable expectation of privacy in a chat room, which was conducted as an open discussion forum in which any adult member of the public could join); *J.S. v. Bethlehem Area Sch. Dist.*, 757 A.2d 412, 425 (Pa. Commw. Ct. 2000) (holding that a student had no reasonable expectation of privacy in a website he created because the website was not password protected); Megan A. Moreno et al., *Research Ethics in the MySpace Era*, 121 (1) PEDIATRICS 157, 158 (2008) (stating that research on MySpace was “analogous to eavesdropping on conversations that take place in a public place such as a coffee shop.”).

expectation of privacy in such tweets.<sup>196</sup> As a matter of fact, Sunlight Foundation's Politwoops was once denied access to Twitter's API on grounds of privacy concerns, though the API access was subsequently restored.<sup>197</sup> Therefore, to avoid a Politwoops-like situation and to comply with Twitter's Privacy Policy, protected tweets and deleted tweets should not be harvested.<sup>198</sup>

Even though there are no privacy concerns related to harvesting publicly available tweets, ethically it may be on murky grounds.<sup>199</sup> This is because even though people may operate in public spaces, they may have a strong perception or expectation of privacy.<sup>200</sup> People may think that, even though the substance of their communication is public, there should be restrictions on how that information is used considering the specific context in which it appears.<sup>201</sup> For

196. Stephen E. Henderson, *Expectations of Privacy in Social Media*, 31 MISS. C. L. REV. 227, 242 (2012) (arguing that there is a reasonable expectation of privacy in once public, now deleted social media content).

197. J.K. Trotter, *Twitter Just Killed Politwoops*, TKTK (June 3, 2015, 5:20 PM), <http://tktk.gawker.com/twitter-just-killed-politwoops-1708842376> (explaining that Politwoops is an application which comprehensively tracks deleted Tweets by public officials). See *Politwoops*, PROPUBLICA, <https://projects.propublica.org/politwoops/> (last visited Dec. 11, 2017) (providing deleted Tweets by public officials). See also Chris Gates, *Eulogy for Politwoops*, SUNLIGHT FOUND. (June 4, 2015, 11:40 AM), <https://sunlightfoundation.com/2015/06/04/eulogy-for-politwoops> (stating that Twitter's decision to pull the plug on Politwoops was a reminder of how the Internet was not truly a public space).

198. Zeynep Tufekci, *Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls*, in PROCEEDINGS OF THE 8TH INTERNATIONAL AAAI CONFERENCE ON BLOGS AND SOCIAL MEDIA (2014) (explaining that there is a very small fraction of private Twitter accounts and that almost all Twitter activity, except for direct messages and private profiles, is visible to the public). See *About Public and Protected Tweets*, TWITTER: HELP DESK <https://help.twitter.com/en/safety-and-security/public-and-protected-tweets#> (last visited Mar. 6, 2019) (explaining that protected Tweets or private Tweets may only be visible to a Twitter user's followers).

199. See *Twitter Terms of Service*, *supra* note 92 (illustrating that even Clause 3 of the Twitter Terms of Service makes it clear that users should only provide content which they are comfortable sharing with others).

200. Annette Markham & Elizabeth Buchanan, *Ethical Decision-Making and Internet Research*, ASS'N INTERNET RESEARCHERS (2002) (showing that while it may be clear that an online newspaper constitutes a public space, the information posted on a publicly accessible Twitter or Facebook profile of an ordinary user might not always be readily defined as public information); Alessandro Acquisti et al., *Privacy and Human Behavior in the Age of Information*, 347 (6221) SCIENCE 509, 513 (2015) (arguing that depending on context, seeking privacy in public is not an oxymoron as individuals can have an expectation of privacy even while sharing information, and even on social media); Malin Sveningsson Elm, *How Do Various Notions of Privacy Influence Decisions in Qualitative Internet Research?*, in INTERNET INQUIRY : CONVERSATIONS ABOUT METHOD 69–87 (Annette Markham & Nancy K. Baym eds., 2009) (arguing that some people might perceive their blogs and Twitter profiles as private spaces or at least as spaces that one would not expect to be observed and analysed for research purposes.); see also Anil Dash, *What is Public?*, THE MESSAGE (July 24, 2014), <https://medium.com/message/what-is-public-f33b16d780f9> (arguing that “[d]on’t publish anything on social media that you wouldn’t want to see on the front page of the newspaper” was an absurd standard since the same tools were being used for interpersonal communications and for making grand pronouncements to the world, often by the same person at different times). See Heidi McKee & James E. Porter, *The Ethics of Digital Writing Research: A Rhetorical Approach*, 59 (4) C. COMPOSITION AND COMM. 1, 27 (2009), [http://www.writing.ucsb.edu/wrconf08/Pdf\\_Articles/McKee\\_Article.pdf](http://www.writing.ucsb.edu/wrconf08/Pdf_Articles/McKee_Article.pdf) (arguing that some participants on the web may perceive their writings on the web as private, rather than published); Andreas Rauber et al., *Ethical Issues in Web Archive Creation and Usage—Towards a Research Agenda*, INT’L WEB ARCHIVING WORKSHOP 3 (2008) (questioning whether sometimes web pages and postings may be considered more like a private communication in public such as a group of friends discussing in a metro train, not meant to be broadcast via the public TV network or to be archived and indexed for everybody to search through).

201. See Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 119, 136 (2004) (expressing this view as “contextual integrity,” which demands that societal information practices should conform to respect context-sensitive privacy norms where context refers not only to overly coarse distinction between “private” and “public,” but to a far richer array of social settings characterized by distinctive roles, norms, and values). For an in-depth insight into contextual integrity, see HELEN NISSENBAUM, *PRIVACY IN CONTEXT* (2009); Markham & Buchanan, *supra* note 200 (arguing that even though people may think that the

example, Twitter users may not be aware that their tweets are available to researchers, even though Twitter's Terms of Service<sup>202</sup> and Privacy Policy<sup>203</sup> is very clear that it has a right to reproduce and distribute the tweets via its API.<sup>204</sup> This is also because people in general are less likely to review online privacy policies and when they do it, many mistakenly believe that the existence of a privacy policy means that their data will not be disclosed or shared with third parties.<sup>205</sup> For this reason, libraries and archives should ask themselves that just

---

substance of their communication is public, there may be restrictions on how that information is used considering the specific context in which it appears.); *see also Social Networking and Ethics*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Aug. 21, 2015), <https://plato.stanford.edu/entries/ethics-social-networking> (arguing that Nissenbaum's view of contextual integrity was well-suited to explain the diversity and complexity of privacy expectations generated by social media.); *Internet Research Ethics*, STAN. ENCYCLOPEDIA PHIL. (June 22, 2012), <https://plato.stanford.edu/entries/ethics-internet-research> (concluding, based on cited sources, that it was unclear whether Internet users truly understood if and when their online activity is regularly monitored and tracked, what kind of reasonable expectations truly exist); THE BRITISH PSYCHOLOGICAL SOCIETY, *Report of the Working Party on Conducting Research on the Internet: Guidelines for Ethical Practice in Psychological Research Online*, 3 (2007), [http://www.bps.org.uk/sites/default/files/documents/conducting\\_research\\_on\\_the\\_internet-guidelines\\_for\\_ethical\\_practice\\_in\\_psychological\\_research\\_online.pdf](http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf) (stating that participants may consider their publicly accessible internet activity to be private, or indeed the communication may even have been private when it first took place); *Research Involving Social Media Data*, U. OF SHEFFIELD: RESEARCH ETHICS POLICY NOTE NO. 14, 1, 4 [hereinafter *Research Involving Social Media Data*], [https://www.sheffield.ac.uk/polopoly\\_fs/1.6709541/file/Research-Ethics-Policy-Note-14.pdf](https://www.sheffield.ac.uk/polopoly_fs/1.6709541/file/Research-Ethics-Policy-Note-14.pdf) (last visited Mar. 6, 2019) (arguing that when assessing the public/private nature of online spaces it is important to take into account that people's perceptions vary and that not all social media users have a good understanding of how accessible their content is to others).

202. *See Twitter Terms of Service*, *supra* note 92 (providing that Clause 3 of Twitter's Terms of Service states that users grant Twitter a worldwide non-exclusive, royalty-free right, to use, copy, reproduce, process, adapt, modify, publish, transmit, display, and distribute such content in any and all media or distribution methods by submitting, posting or displaying content on Twitter).

203. *See* TWITTER, <https://twitter.com/en/privacy> (last visited Mar. 6, 2019) (providing that Twitter's Privacy Policy states that it has the right to share tweets, retweets, or any other information shared with it).

204. *See* Sara Mannheimer, *Improving Services - At What Cost? Examining the Ethics of Twitter Research at the Montana State University Library*, COUNCIL FOR BIG DATA, ETHICS, & SOCIETY 8 (2016), <http://bdes.datasociety.net/wp-content/uploads/2016/10/MannheimerYoungRossmann-twitterlibrary.pdf> (arguing that Twitter users may not be aware that their data is being made available to researchers through the Twitter API). Survey respondents have expressed that the long-term storage of tweets has never been discussed by the founders. Further, while Twitter's Terms of Service may grant Twitter the legal right to archive tweets in perpetuity, there was no language in the Terms of Service indicating that this was actually happening. Nicholas Proferes, *What Happens to Tweets? Descriptions of Temporality in Twitter's Organizational Rhetoric*, ICONFERENCE PROCEEDINGS 76, 82 (2014) [https://www.ideals.illinois.edu/bitstream/handle/2142/47313/045\\_ready.pdf?sequence=2](https://www.ideals.illinois.edu/bitstream/handle/2142/47313/045_ready.pdf?sequence=2); *see, e.g.,* Michael Zimmer, *Is it Ethical to Harvest Public Twitter Accounts without Consent?*, MICHAEL ZIMMER (Feb. 12, 2010), <http://www.michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent> (arguing that just because a Twitter user has made the tweets visible does not mean automatic consent that the tweet can be harvested, archived and mined by researchers). A survey respondent has expressed a sense of uncertainty about what goes online. Catherine C. Marshall & Frank M. Shipman, *Social Media Ownership: Using Twitter as a Window Onto Current Attitudes and Beliefs*, SIGCHI CONF. ON HUM. FACTORS IN COMPUTING SYS. 1081, 1085 (2011), <https://www.microsoft.com/en-us/research/wp-content/uploads/2011/05/chi2011-Marshall-Shipman.pdf> (discussing a respondent stating "I don't usually publish too much information about myself because I don't know who is going to look at it."); *Twitter Terms of Service*, *supra* note 92 (stating that users grant Twitter "a worldwide non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods" by submitting, posting or displaying content on Twitter).

205. *See* HARRY EVANS ET AL., IPSOS MORI, #SOCIALETHICS: A GUIDE TO EMBEDDING ETHICS IN SOCIAL MEDIA RESEARCH 31 (2015), <https://www.ipsos.com/sites/default/files/publication/1970-01/im-demos-social-ethics-in-social-media-research.pdf> (finding that only 38% of survey respondents were aware that their social media data was available to third parties such as the government or companies for research purposes); Nathaniel S. Good et al., *Noticing Notice: A Large Scale Experiment on the Timing of Software License Agreements*, CHI 1 (2007), <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.8044&rep=rep1&type=pdf> (stating that



because they can technically and legally archive tweets, should they actually do it?<sup>206</sup> And if yes, can they subsequently publish it and make it available to

---

most users do not bother to read the lengthy and legalistic End User License Agreements (EULAs) or Privacy Agreements); Luke Hutton & Tristan Henderson, "I Didn't Sign Up for This!" *Informed Consent in Social Network Research*, NINTH INT'L AAA CONF. ON WEB & SOC. MEDIA 178, 178–79 (2015), <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10493/10501> (arguing that while users may have formally agreed to their data being used by accepting the social media platform's Terms of Service, they may not be aware of being actually observed by researchers); Tony Vila et al., *Why We Can't Be Bothered to Read Privacy Policies: Models of Privacy Economics as a Lemons Market*, MIT (May 29, 2003) [http://web.mit.edu/~greenie/Public/weis\\_slides.pdf](http://web.mit.edu/~greenie/Public/weis_slides.pdf) (stating that consumers do not read privacy policies); Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. CHI. L. REV. 1203, 1217 (2003) (stating that over the years, commentators have routinely observed that buyers often fail to read the terms of standard form contracts); Yannis Bakos et al., *Does Anyone Read the Fine Print? Consumer Attention to Standard Form Contracts* 3 (N.Y. Univ. Law & Econ. Working Paper No. 195, 2014), [http://lsr.nellco.org/nyu\\_lewp/195](http://lsr.nellco.org/nyu_lewp/195) (stating that a small fraction of consumers read standard form contracts or choose to be informed about standard form online contracts); Robert A. Hillman, *Online Consumer Standard Form Contracting Practices: A Survey and Discussion of Legal Implications* 1–2 (Cornell Law Sch. Research Paper No. 05-012, 2005), [http://scholarship.law.cornell.edu/cgi/viewcontent.cgi?article=1028&context=lsrp\\_papers](http://scholarship.law.cornell.edu/cgi/viewcontent.cgi?article=1028&context=lsrp_papers) (stating that people in general do not read their e-standard forms.); Joseph Turow et al., *Open to Exploitation: America's Shoppers Online and Offline* 20 (Annenberg Public Policy Center of the University of Pennsylvania, Working Paper No. 35, 2005), [http://repository.upenn.edu/asc\\_papers/35](http://repository.upenn.edu/asc_papers/35) (finding that fifty-nine per cent of survey respondents believed that when a website had a privacy policy it meant that it would not share information with other websites or companies). Cf. Aaron Smith, *What Internet Users Know About Technology and the Web*, PEW RES. CTR. (Nov. 25, 2014), <http://www.pewinternet.org/2014/11/25/web-iq> (finding that forty-four per cent of survey respondents believed that when a company posts a privacy statement, it did not necessarily mean that they were actually keeping user information confidential).

Further, even if users read privacy policies, they often fail to make informed decisions. Alessandro Acquisti & Jens Grossklags, *Privacy and Rationality in Individual Decision Making*, 3 IEEE SEC. & PRIVACY 24, 26 (2005), <https://www.dtc.umn.edu/weis2004/acquisti.pdf> (finding inconsistency in privacy decision making and behavior); Carlos Jensen et al., *Privacy Practices of Internet Users: Self-Reports Versus Observed Behavior*, 63 INT'L. J. HUM.-COMPUTER STUD. 203, 226 (2005), <https://www.sciencedirect.com/science/article/abs/pii/S1071581905000650?via%3Dihub> (stating that not only do users frequently fail to consult online privacy policies but when they do, it may not help them to make informed decisions); see also Acquisti et al., *supra* note 200, at 511 (stating that even though people may be aware of the consequences of their privacy decisions, they are likely to be unaware about their own privacy preferences); Lauren Dugan, *Fifty-Two Percent of Twitter Users Do Not Consider Legal Implications of Their Tweets*, ADWEEK (Nov. 3, 2011), <http://www.adweek.com/digital/twitter-legal/?red=at> (stating that only 18% of social media users surveyed in 2011 had read the terms and conditions for posting to the social media sites they used, a decline compared to 33% in 2008); Carlos Jensen & Colin Potts, *Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices*, 6 SIGCHI CONF. ON HUM. FACTORS IN COMPUTING 471, 477 (2004), <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.629.1633&rep=rep1&type=pdf> (stating that even though users were concerned about their privacy and claim to take steps to protect themselves, they tend to grossly "over-report the frequency and likelihood with which they read privacy policies."); Patricia A. Norberg et al., *The Privacy Paradox: Personal Information Disclosure Intentions Versus Behaviors*, 41 J. CONSUMER AFF. 100, 101 (2007) (concluding that people had a paradoxical attitude towards privacy where they intended to protect their privacy but also provide personal details); Sarah Spiekermann et al., *E-Privacy in 2nd Generation E-Commerce: Privacy Preferences Versus Actual Behavior*, 48 CACM 38, 45 (2005), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=761107](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=761107) (indicating a surprising readiness on the part of online users to reveal their private and even highly personal information).

206. Sara Mannheimer et al., *Improving Services-At What Cost? Examining the Ethics of Twitter Research at the Montana State University Library*, THE COUNCIL FOR BIG DATA, ETHICS & SOC'Y 10 (Oct. 2016), <http://bdes.datasociety.net/wp-content/uploads/2016/10/MannheimerYoungRossmann-twitterlibrary.pdf>; see also Nicholas Taylor, *Questions of Ethics at Web Archives 2015*, STAN. LIBR. (Dec. 17, 2015), <http://library.stanford.edu/blogs/digital-library-blog/2015/12/questions-ethics-web-archives-2015> (terming this as the "problematization of a binaristic conception of privacy" which means that publicness of social media content is taken as tacit approval to archive and redistribute the content); Yvonne Ng, *Ethical Wednesdays: Archives and Our Ethical Guidelines for Using Eyewitness Videos*, WITNESS MEDIA LAB (Nov. 2015), <https://lab.witness.org/ethical-wednesdays-archives-and-our-ethical-guidelines-for-using-eyewitness-videos> (questioning whether, on ethical grounds, archives should collect tweets).

researchers without seeking the consent of the concerned Twitter user?<sup>207</sup> This becomes specially relevant considering that some Twitter users feel strongly that they should be able to store and archive their tweets, rather than a third party.<sup>208</sup>

During my time on the SFM project, the SFM team delved into this issue at length and ruled out a permissive approach. After extensive deliberations, there was unanimity that as a university library with limited staff, it was impractical to seek the consent of each and every Twitter user before harvesting their tweets.<sup>209</sup> However, to comply with Twitter's developer policy and to be on ethically firm ground, it was decided to capture the metadata of the archived tweets.<sup>210</sup> Further, it was also decided to notify the researchers about the following:

---

207. Melissa Bica & Jennings Anderson, "You Are What You Tweet!" *The Ethics of (Re) Publishing Public Data as Crafted Narratives*, ACM CONF. ON HUM. FACTORS IN COMPUTING SYSTEMS (2016), <https://ethicalencountershci.files.wordpress.com/2016/03/bica-and-anderson.pdf>. To honor a Twitter user's reasonable expectation of privacy, when the Center for Media & Social Impact collected Twitter data documenting the Black Lives Matter movement, they only published tweets which had been widely shared and had a large number of followers. See DEEN FREELON ET AL., CTR. FOR MEDIA & SOC. IMPACT, BEYOND THE HASHTAGS 86 (2016), [http://cmsimpact.org/wp-content/uploads/2016/03/beyond\\_the\\_hashtags\\_2016.pdf](http://cmsimpact.org/wp-content/uploads/2016/03/beyond_the_hashtags_2016.pdf) (stating that several steps were taken to protect the privacy and intellectual property of the Twitter users whose usernames and tweets were collected); MARK S. FRANKEL & SANYIN SIANG, AM. ASS'N FOR THE ADVANCEMENT OF SCI., ETHICAL AND LEGAL ASPECTS OF HUMAN SUBJECT RESEARCH ON THE INTERNET 1, 7 (1999) [https://www.researchgate.net/publication/268296040\\_Ethical\\_and\\_Legal\\_Aspects\\_of\\_Human\\_Subjects\\_Research\\_on\\_the\\_Internet](https://www.researchgate.net/publication/268296040_Ethical_and_Legal_Aspects_of_Human_Subjects_Research_on_the_Internet) (arguing that three features of the internet, the blurred distinction between the private vs. public domain, its easy conductivity for anonymous and pseudonymous communications, and its global and easy accessibility, posed difficult questions for informed consent).

208. Marshall & Shipman, *supra* note 204, at 1085; see also EVANS ET AL., *supra* note 205, at 31 (finding 60% of survey respondents felt that their social media data should not be shared with third parties, such as the government or companies, for research purposes).

209. See Matthew Zook et al., *Ten Simple Rules for Responsible Big Data Research*, 13 PLOS COMPUTATIONAL BIOLOGY (Mar. 30, 2017), <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005399> (arguing that when dealing with large volumes of user generated content, obtaining informed consent from the users of a social media platform is difficult or even impossible). Considerable discussion has taken place in internet ethics literature regarding the need for informed consent, where it is accepted that the majority of web sites, discussion boards, and chat rooms cannot be considered as private spaces but as public spaces. Mary R. Weeden, *Ethics and On-Line Research Methodology*, 9 J. SOC. WORK VALUES & ETHICS 40, 43 (2012); see also Adolfo Estalella & Elisenda Ardèvol, *Field Ethics: Towards An Ethic Located for the Ethnographic Research of Internet*, 8 FORUM: QUALITATIVE SOC. RES. (2007) (arguing that despite the importance given to informed consent, ethical guidelines propose an exception for when the research is carried out in public spaces, extending to social media such as the press, the radio or television); *Research Involving Social Media Data*, *supra* note 201, at 6 ("As with all research involving observation of public space it is recognised that it is often infeasible and unnecessary" to seek the consent of all who may be observed). Also, though this point never came during my time on the SFM project, it may be particularly challenging to seek consent of Twitter users as they may not fully understand the true objective of the project. See David Pittenger, *Internet Research: An Opportunity to Revisit Classic Ethical Problems in Behavioral Research*, 13 ETHICS & BEHAV. 45, 58 (2003) (arguing that the primary problem is that internet does not guarantee that the individual supplying informed consent understand the ramifications of consent or is qualified to give consent); Emily Wolfinger, "But its already public, right?": *The Ethics of Using Online Data*, DATA DRIVEN JOURNALISM (Nov. 25, 2016), [http://datadrivenjournalism.net/news\\_and\\_analysis/but\\_its\\_already\\_public\\_right\\_the\\_ethics\\_of\\_using\\_online\\_data](http://datadrivenjournalism.net/news_and_analysis/but_its_already_public_right_the_ethics_of_using_online_data) (discussing the challenges of obtaining informed consent from users). Further, there is no guarantee that the Twitter users will read the informed consent form carefully.

210. *Display Requirements: Tweets*, TWITTER [hereinafter *Display Requirements*], <https://developer.twitter.com/en/developer-terms/display-requirements> (last visited Mar. 6, 2019) (stating that the "Tweet author's profile picture, @username, and display name must always be displayed and link[ed] to the user's Twitter profile.").

- (i) That the collection was only available for research purposes.
- (ii) Any form of commercial usage was strictly prohibited.<sup>211</sup>
- (iii) Substantial reproduction, redistribution or bulk download of the collection was strictly prohibited.<sup>212</sup>
- (iv) The name of the Twitter user(s) whose data set is being reproduced should be anonymized in research papers and other publications.<sup>213</sup>
- (v) The researcher had the onus to comply with any copyright restrictions.

Based on an analysis of the access policies of similar archives such as the U.S.-Cuba Policy Change Twitter archive at the University of Miami and the Tweeting#OWS Project at the Emory University, there was also a consensus that access to the collection may have to be restricted even though restricting access severely undermined Gelman Library's mandate to maximize accessibility to archival materials.<sup>214</sup> As a matter of fact, all libraries and archives are duty bound to ensure widest possible accessibility to materials.<sup>215</sup> Keeping this in

---

211. This is to comply with Twitter's Developer Policy which states that tweets should not be used to promote any product or service for example by displaying the tweets in advertising or otherwise displaying them to imply endorsement of any product or service without the user's explicit permission. *Id.*

212. This is to comply with Twitter's Developer Policy which states that no redistribution of tweets was permissible except as explicitly mentioned. See *Developer Agreement and Policy*, TWITTER, <https://developer.twitter.com/en/developer-terms/agreement-and-policy> (last visited Mar. 6, 2017). In the past, Twitter has requested institutional authors to remove datasets from their institutions website so that it could not be freely downloaded. See Katharina Kinder-Kurlanda et al., *Archiving Information from Geotagged Tweets to Promote Reproducibility and Comparability in Social Media Research*, 4 *BIG DATA & SOC'Y* 1, 5 (2017), <http://journals.sagepub.com/doi/pdf/10.1177/2053951717736336>. Further, the Twitter Project Page at the Max Planck Institute for Software Systems explicitly states that it is not allowed to share any tweet information. *The Twitter Project Page at MPI-SWS*, THE TWITTER PROJECT PAGE AT MPI-SWS, <http://twitter.mpi-sws.org> (last visited Mar. 6, 2019).

213. See *Display Requirements*, *supra* note 210 (stating that the "Tweet author's profile picture, @username, and display name must always be displayed and link[ed] to the user's Twitter profile."); *Research Involving Social Media Data*, *supra* note 201, at 6 (proposing that steps should be undertaken to anonymize social media users in publications and other outputs, unless the individual is a public figure acting in a public capacity); TOWNSEND & CLAIRE WALLACE, *supra* note 195 (arguing that the issue of anonymization becomes more critical in cases where data sets or individual units of data are published online, in academic journals, and at academic conferences, and when the data pertains to sensitive subject matter).

214. See Natalie Baur, *2014 US-Cuba Policy Change Twitter Archive*, U. MIAMI SCHOLARLY REPOSITORY (2015), [https://scholarlyrepository.miami.edu/chc\\_archiveddatasets/1](https://scholarlyrepository.miami.edu/chc_archiveddatasets/1) (restricting access to the US-Cuba Policy Change Twitter Archive of the University of Miami Cuban Heritage Collection to the University of Miami community via an IP-restricted portal in the University of Miami Scholarly Repository). The SFM team was told by Katie Rawson, in-charge of the Tweeting#OWS project designed by the Emory Libraries' Digital Scholarship Commons (DiSC), that access to the archive was restricted only to scholars affiliated with the Emory University. See Karishma Mehrotra, *Library Archives Protest Tweets*, EMORY WHEEL (Sept. 25, 2012), <http://emorywheel.com/library-archives-protest-tweets> (describing the Tweeting#OWS project); see also Leslie King, *Emory Digital Scholars Archive Occupy Wall Street Tweets*, EMORY NEWS CTR. (Sept. 21, 2012), [http://news.emory.edu/stories/2012/09/er\\_occupy\\_wall\\_street\\_tweets\\_archive/campus.html](http://news.emory.edu/stories/2012/09/er_occupy_wall_street_tweets_archive/campus.html) (describing the Tweeting#OWS project).

215. See *SAA Core Values Statement and Code of Ethics*, SOC'Y AM. ARCHIVISTS, [https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics#code\\_of\\_ethics](https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics#code_of_ethics) (last visited Mar. 6, 2019) (advocating the widest possible accessibility of materials, consistent with mandatory access restrictions such as public statute, donor contract, business/institutional privacy, or personal privacy); INT'L COUNCIL ON ARCHIVES, *CODE OF ETHICS 2*, [https://www.ica.org/sites/default/files/ICA\\_1996-09-06\\_code%20of%20ethics\\_EN.pdf](https://www.ica.org/sites/default/files/ICA_1996-09-06_code%20of%20ethics_EN.pdf) (last visited Mar. 16, 2019) (advocating that archivists should discourage unreasonable access restrictions).

mind, and to comply with Twitter's Privacy Policy, we concluded that the collection may be accessible to all scholars irrespective of their university affiliation in the reading room of the Special Collections Research Center.<sup>216</sup> However, remote access via user authentication should only be available to G.W. affiliated scholars.<sup>217</sup> Further, to comply with Twitter's Developer Policy, it was decided not to harvest deleted tweets through a technical process known as rehydration (also known as hydration).<sup>218</sup> With this technique, it is hoped that privacy concerns of Twitter users are partially ameliorated. Closely related to the right of privacy is the right of publicity, which libraries and archives should be mindful of before creating a Twitter archive.<sup>219</sup>

#### F. Right of Publicity

The right of publicity prohibits the usage of an individual's "name, image or likeness" for a "commercial purpose"<sup>220</sup> unless proper consent of the individual whose name, image or likeness is being used has been sought.<sup>221</sup> The right usually vests in celebrities though the majority view is that it may extend to every individual and not just celebrities. It is protected primarily by state

216. See THE SECTION 108 STUDY GROUP, *supra* note 106, at viii (recommending that once the publicly available online content is captured, access to researchers should be provided on the library's or archives' premises).

217. See *id.* (following the recommendations of the Section 108 Study Group, remote access should only be provided after a lapse of a reasonable time period).

218. See *Developer Policy*, TWITTER, <https://developer.twitter.com/en/developer-terms/policy#ii-rules-for-specific-twitter-services-or-features> (last visited Mar. 16, 2019) (stating that if the "Content is deleted, gains protected status, or is otherwise suspended, withheld, modified, or removed from the Twitter Service," reasonable effort should be made to delete or modify such content as soon as possible or within 24 hours after a request to do so is made by Twitter or the user); see also *Research Involving Social Media Data*, *supra* note 201, at 7 (stating that a significant issue in social media research was how to handle deleted posts). Rehydration is the technique of un-duplicating data. In the context of SFM, once a tweet has been originally harvested by SFM, to rehydrate it would mean to bring the export at the time of delivery to its original state, i.e., the tweets which were subsequently deleted will no longer be a part of the final extract which is made available to the researcher. See generally Kinder-Kurlanda et al., *supra* note 212, at 5 (explaining the rehydration technique); see also *On Forgetting*, INKDROID (Nov. 18, 2014), <https://inkdroid.org/2014/11/18/on-forgetting> (describing the hydration technique).

219. See *Privacy and Publicity Rights*, LIBR. CONG., <http://lcweb2.loc.gov/ammem/copothr.html> (last visited Mar. 16, 2019) (discussing the difference between privacy rights and publicity rights from the perspective of libraries and archives).

220. Floyd A. Gibson & Rachel M. Healey, *The Right of Publicity Comes of Age*, 23 AIPLA Q. J. 361, 375 (1995).

221. Sudakshina Sen, *Fluency of the Flesh: Perils of an Expanding Right of Publicity*, 59 ALB. L. REV. 739, 742 n.20, 744 n.40 (1996). See generally K. J. Greene, *Intellectual Property Expansion: The Good, the Bad, and the Right of Publicity*, 11 CHAP. L. REV. 521, 528–33 (2008) (describing the rationales for protecting the right of publicity); Mark S. Lee, *Agents of Chaos: Judicial Confusion in Defining the Right of Publicity-Free Speech Interface*, 23 LOY. L.A. ENT. L. REV. 471, 478 (detailing a brief history of the right of publicity in the United States); *A Brief History of the Right of Publicity*, RIGHT OF PUBLICITY (July 31, 2015), <http://rightofpublicity.com/brief-history-of-rop> (discussing the rationale for the right to publicity and the legal requirement of consent); Sudakshina Sen, *Fluency of the Flesh: Perils of an Expanding Right of Publicity*, 59 ALB. L. REV. 739, 739–40 (1996) (detailing the rationale for protecting the right of publicity); Fred M. Weiler, *The Right of Publicity Gone Wrong: A Case for Privileged Appropriation of Identity*, 13 CARDOZO ARTS & ENT. L.J. 223, 240–45 (1994) (describing the rationales for protecting the right of publicity); Stephen M. Lobbin, *The Right(s) of Publicity in California: Is Three Really Greater Than One?*, 2 UCLA ENT. L. REV. 157, 174–178 (1995) (discussing the right of publicity after death).

law.<sup>222</sup> Though public libraries are not typically engaged in “commerce,”<sup>223</sup> this right is relevant with respect to a Twitter archive in two situations:

1. *If Celebrity’s Image is Used for Advertising or Promoting the Twitter Archive*

Depending upon a state’s right of publicity law, a library or archive should under most circumstances refrain from using a celebrity’s name, image or likeness, without the celebrity’s consent: (i) for identifying the proprietorship in the Twitter archive, (ii) for the purpose of advertising or soliciting patrons to the Twitter archive, (iii) for raising funds for financing the Twitter archive.<sup>224</sup> These are just some general restrictions based on Ohio’s law of publicity. For specific restrictions, libraries and archives should consult their jurisdiction’s publicity law before using a celebrity’s persona for commercial purposes.

2. *If a Tweet Related to a Celebrity is Harvested.*

If a library or archive harvests a tweet related to a celebrity, it is ideally the Twitter user who should seek the celebrity’s permission especially if it is being tweeted or posted in a commercial context.<sup>225</sup> However, to be on a firmer footing, libraries and archives should seek the consent of the celebrity before using the harvested tweet for: (i) identifying the proprietorship in the Twitter archive (ii) for the purpose of advertising or soliciting patrons to the Twitter archive (iii) for raising funds for financing the Twitter archive so as to give a false impression of endorsement.<sup>226</sup>

The *sine qua non* of a right of publicity claim is the use of a celebrity’s “name, image or likeness” for commercial purposes.<sup>227</sup> Therefore, a non-commercial Twitter archive which aids research and scholarship is unlikely to violate the right of publicity of a celebrity.<sup>228</sup>

---

222. See *Statutes and Interactive Map*, RIGHT OF PUBLICITY, <http://rightofpublicity.com/statutes> (last visited Mar. 16, 2019) (listing state wise right of publicity statutes).

223. See *White v Samsung Elec. Am., Inc.*, 971 F.2d 1395, 1401 (9th Cir. 1992) (holding that the use of celebrity’s name in a parody did not necessarily render the use non-commercial.); PETER B. HIRTLE ET AL., COPYRIGHT AND CULTURAL INSTITUTIONS 179 (2009) (stating that right of publicity “should not apply to non-commercial, educational use of a person’s image”); *Implications of Right of Publicity on Library Activities*, OHIO LIBR. COUNCIL (June 11, 2009) [hereinafter *Implications of Right of Publicity on Library Activities*], <http://www.olc.org/pdf/VorysRightOfPublicityLibraryActivities112408.pdf> (stating that though public libraries are not typically engaged in commerce, “libraries should assume that for the right of publicity purposes, their uses of ‘persona’ is commercial in nature”).

224. See *Implications of Right of Publicity on Library Activities*, *supra* note 223 (noting the restrictions are based on Ohio’s law of publicity).

225. Squire Patton Boggs, *Celebrity Tweets Can Cost You Millions*, SQUIRE PATTON BOGGS: GLOBAL IP & TECH. L. BLOG (June 2, 2014), <https://www.iptechblog.com/2014/06/celebrity-tweets-can-cost-you-millions>.

226. This is based on Ohio’s law of publicity and would differ from state to state. See *Implications of Right of Publicity on Library Activities*, *supra* note 223 (applying Ohio’s law of publicity).

227. See *Gibson & Healey*, *supra* note 217, at 375 (finding that “to sustain a claim of a right of publicity, the plaintiff must prove that the taking of his name, [image or] likeness was in fact for a commercial purpose and not merely incidental use.”).

228. See *Rogers v. Grimaldi*, 875 F.2d 994, 1005 (2d Cir. 1989) (holding that non-commercial use of a celebrity’s name in a film title was immune from right of publicity assertions); *Paulsen v. Personality Posters, Inc.*, 299 N.Y.S.2d 501, 506 (N.Y. 1968) (“The privilege of enlightening the public is by no means limited to

#### IV. CONCLUSION

The primary aim of this Article has been to inform the libraries and archives community involved in archiving tweets about the potential legal issues that they may encounter during the course of the project, which are also applicable to general social media archiving. At a broader level, it argues that the legal framework for creating a social media archive is strong. However, ethical issues specifically with respect to user privacy persist. Even though conversations on ethical issues have already started to take place at leading libraries and archives conferences, there is an urgent need for calibrated effort to document the ethical dimensions of social media archiving by the stakeholders concerned.

---

dissemination of news in the sense of current events but extends far beyond to include all types of factual, educational and historical data, or even entertainment and amusement, concerning interesting phases of human activity in general.”). See generally Andrew Koo, *Right of Publicity: The Right of Publicity Fair Use Doctrine - Adopting a Better Standard*, 4 BUFF. INTELL. PROP. L.J. 1, 21 (2006) (considering a fair-use analysis of right of publicity claims and proposing that a right of publicity fair use doctrine be adopted to properly balance First Amendment rights and the right of publicity).