

Winnie Schwaid-Lindner

Intro to MIAP

Due December 8, 2016

Howard Besser

Archival Ethics and Law of Social Media

Introduction

Archives, as memory institutions, are often designed around access. However, in the past 15 years with the development of social media platforms, the factors that inhibit access have become more complex than ever before. As difficult as it may be to provide access to a book or film from 1965, when you enter the realm of social media, the intricacies triple. For each specific medium, the laws, ethics, and privacy of these platforms must be taken into account. Each individual social media platform, website, and digital publication possesses its own set of internal rules and regulations. The reconciliation of these with laws and ethics are what will permit us to integrate these materials and provide access to them. To consider this, we must also address that each specific medium was created with a singular, yet differing intention.

For an example that will surely date itself in the years to come, Twitter and Snapchat are both social medias, and yet one is text-based, and the other visual. The former has very limited privacy modifications, the latter is almost exclusively used for private communication. Twitter has established a precedent by sharing all of its past 'tweets' with the Library of Congress, and has been at the forefront of data exchange, while Snapchat maintains that it saves no user interactions, maintaining the highest

level of security and privacy. What are the legal and ethical limitations of trying to put these similarly contradictory mediums into a singular archival setting? Here, we will consider these implications, as discussed by several experts throughout the fields of data science, law, and information studies in order to best determine how to balance new media in an real world archival setting.

Legality

Delving first into the legal implications of new media archiving, there are copyright and privacy laws that may factor into providing access of archived materials. Regarding copyright laws specifically, there is the question of whether this material is protected at all. In order to meet the threshold standard for copyright protection, a work must exhibit a “modicum of creativity”¹. Although many longer posts clearly qualify, there is the matter of shorter posts. It does not necessarily take creativity to share an image accompanied with a small caption, nor to write a one sentence post about your day. Due to this, there is an argument to be made that many social media posts are not protected at all.

In addition to the modicum of creativity, works that are broadly published do not qualify for copyright protection. As the case of *Estate of Martin Luther King, Jr., Inc. v. CBS, Inc.* notes: “A general publication occurred ‘when a work was made available to members of the public at large without regard to their identity or what they intended to

¹ “FEIST PUBLICATIONS, INC. v. RURAL TELEPHONE SERVICE CO., 499 U.S. 340 (1991).” Accessed December 3, 2016. https://www.law.cornell.edu/copyright/cases/499_US_340.htm.

do with the work”². When celebrities with more than 2 million followers post a tweet, are they monitoring who receives this content and noting their followers’ intentions?

Furthermore, the Copyright Office notes that “short phrases” are not eligible for copyright protection. Unfortunately, there is no specification for what a “short phrase” entails. The case of *CRA Mktg., Inc. v. Brandow’s Fairway Chrysler-Plymouth-Jeep-Eagle, Inc.* indicates that 54 words is enough to garner copyright protection³, however with a 140 character maximum, the usual tweet is a far cry from 54 words. For example, the previous sentence (The case ... words.) clocks in at 40 words and 230 characters - nearly 100 more than Twitter permits.

As an archival student, I do not hold the qualifications to determine conclusively whether these two circumstances would warrant protection or not. However, even discounting the content that may be devoid of creativity and broadly published, there are still billions of social media posts that do qualify for copyright protection. For each of these posts, there is a corresponding Terms of Service issued by the platform that indicates that they can “...use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods (now known or later developed)”⁴. What this means in practical terms for an archive is that if you can establish a license with the platform (as the Library of

² Anderson, Chief Judge, Senior Circuit Judge Roney, and Senior District Judge Cook. “Estate of Martin Luther King, Jr., Inc. v. CBS, Inc.,” n.d.

³ Minow, Mary. “Copyright Protection for Short Phrases – Rich Stim.” *Stanford Copyright and Fair Use Center*, September 9, 2003.

http://fairuse.stanford.edu/2003/09/09/copyright_protection_for_short/.

⁴ “Terms of Service.” *Twitter*. Accessed December 3, 2016. <https://twitter.com/tos>.

Congress has done with Twitter), it is entirely legal to archive these materials, and distribute it as widely as possible within the terms of the license.

Furthermore, within the realm of copyright there is the possibility of a Fair Use claim. In order to prove fair use, we can evaluate that the use of archiving social media materials would be transformative based on the use of the archive being non-social or entertainment driven (which is the original use of social media). In addition, the precedent of *Authors Guild v. Google, Inc.*, which indicates that the data and cumulative nature of the content (an archive would be taking all social media content from all users in an unbiased fashion) acquired from archiving these materials is a fair use⁵.

In terms of other laws that may affect the preservation of social media, there are laws designed to protect an individual's privacy. However, while there are laws that regulate the use of an individual's social media in regards to employers and educational institutions, these are designed to protect from the use of a social media presence to ascertain the character of an applicant or employee. These regulations are dictated on a statewide basis, with a total of 15 states having legislation that protects against the use of personal social media by any type of institution⁶. As this law often protects specifically against requiring the employee/student to provide their own information⁷, if the essence

⁵ "Authors Guild v. Google, Inc., No. 13-4829 (2d Cir. 2015)." *Justia Law*. Accessed December 3, 2016.

<http://law.justia.com/cases/federal/appellate-courts/ca2/13-4829/13-4829-2015-10-16.html>.

⁶ "STATE SOCIAL MEDIA PRIVACY LAWS," July 6, 2016.

<http://www.ncsl.org/research/telecommunications-and-information-technology/state-laws-prohibiting-access-to-social-media-username-and-passwords.aspx>

⁷ MacFARLAND, Millicent M. "An Act To Protect Social Media Privacy in School and the Workplace H.P. 838 No. 1194," March 26, 2013.

in particular were not acquired through one of these manners and was found legally, then there's no law prohibiting it.

The strictest law about privacy in regards to electronic communication is that which concerns the distribution of pornographic materials without consent. This is known colloquially as "Revenge Pornography", and since it is most commonly used for personal gain through revenge or humiliation, the law is only regarding the specific publication or distribution of such materials, and therefore this law doesn't reach the jurisdiction of casual social media usage⁸. Likewise, state-governed Right of Publicity laws come close to covering this issue, but stop short. The Right of Publicity only covers the financial benefits of using one's image or likeness. As a memory institution, it is unclear how much commercial or fiscal benefit will really come from saving or providing access to tweets. Furthermore, considering the sheer number of individual posts involved, it would be especially unlikely that any singular tweet would provide a significant financial incentive.

Ethics

Although it may be legal to archive new media materials, the law can often veer from ethics. Simply because it is possible for us to do so, doesn't mean that it is necessarily the responsible course of action. Anticipating deliberations by archivists, the SAA (Society of American Archivists) has written a code of ethics. This code does speak to both the matter of both privacy and access. It makes the point that "archivists

⁸ Citron, Danielle, and Mary Franks. "Criminalizing Revenge Porn." *Faculty Scholarship*, January 1, 2014. http://digitalcommons.law.umaryland.edu/fac_pubs/1420.

place access restrictions on collections to ensure that privacy and confidentiality are maintained, particularly for individuals and groups who have no voice or role in collections' creation, retention, or public use" and continues to note that "in all questions of access, archivists seek practical solutions that balance competing principles and interests."⁹ Based on these guidelines, it's possible to evaluate the ethical standards for archiving social media materials.

One of the most fundamental credences of a research scientist is that one's research must not harm the subject or community of which you're researching. In many archival uses, this will come down to access. In 2008, a study was conducted called the "Taste, Ties, and Time" project. Sensitive materials regarding this project were discovered by third parties against the wishes and knowledge of the researchers and subjects, leading to numerous threats and the general lack of safety that should be afforded as a human right¹⁰. There are many issues that can come from private information being released digitally. But from an archival standpoint, if you were to collect the information and not release it (jeopardizing the safety of the authors), then is there an ethical dilemma at all? In fact, Zimmer notes that "...recognizing that just because personal information is made available in some fashion on a social network, does not mean it is fair game for capture and release to all." This quote specifically mentions "capture and release to all", indicating that perhaps, it is the capture *and*

⁹ "SAA Core Values Statement and Code of Ethics | Society of American Archivists." Accessed December 3, 2016.

<http://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics>.

¹⁰ Zimmer, Michael, zimmerm@uwm.edu. "But the Data Is Already Public': on the Ethics of Research in Facebook." *Ethics & Information Technology* 12, no. 4 (December 2010): 313–25. doi:10.1007/s10676-010-9227-5.

release of this data that is unethical and causes harm. It is not necessarily one entirely at the expense of the other. There are means to provide access to this information while still protecting the participants of the archive on an individual level. An access focused strategy would be to obscure the personal information associated with the posts. This way, all of the original content is provided, yet none of the potentially harmful information is released. A more stringent and privacy-focused method may be to release the data findings, but not to release the original content itself. An example of this may look like a search tool, similar to Google Ngrams or Google Trends. In this scenario, one could search for terms, and see the key's usage over time. The search key "Earthquake Japan" would display the usage of these terms with a spike indicating more posts, and these spikes would correlate to the day of an earthquake in Japan. This strategy would render this data incredibly useful, but not ethically questionable.

On the anthropological side, assuming that we have the ability to acquire and implement this information, is it ethical for us to do so without the knowledge of the author? If the author *were* to have knowledge that their data was being taken, the distinction between full comprehension of the provision of their work (also known as informed consent¹¹) and a superficial knowledge of simply handing their work over, also is worth discussion. As Ilka Gleibs notes:

"New challenges and questions arise when we deal with data ... for example, who is responsible for making sure individuals and communities are not hurt by

¹¹ "Ethics Task Force – Draft Principle: Informed Consent." *Welcome to the AAA Blog*, March 3, 2011. <https://blog.americananthro.org/ethics-task-force/ethics-task-force-draft-principle-informed-consent/>.

the research? What does it mean for ... users to provide behavioral data without their knowledge and to what extent are their rights violated? On a practical level, what does consent (if needed) look like when using data...?"¹²

It is undeniable that asking every author whether each of their works can be archived is entirely unfeasible. There are currently over 1.7 billion active users on Facebook alone¹³, and if we were to archive each of their posts, it would quickly become unmanageable in terms of both time and sheer scope of work. Luckily, this may be where Terms of Service and licensing agreements can cover some ground. Although the aforementioned documents are a legal hurdle, they provide some clarity when considering ethics.

If one's rights are signed over initially in the Privacy Policy and/or the Terms and Conditions section of the website, then that nebulous legal jargon may inherently mean that it was never in any individual's power to keep our social media private in the first place. As an example, Twitter reserves all rights to share any public information. Although you maintain the rights to all your individual content, the data associated with your media belongs to the site itself¹⁴. This is usually used for advertising and other third parties, but it also indicates that there will be no conflict for saving the metadata associated with the essence, regardless of the ethical conflict of the media itself. In

¹² Gleibs, Ilka H. "Turning Virtual Public Spaces into Laboratories: Thoughts on Conducting Online Field Studies Using Social Network Sites." *Analyses of Social Issues and Public Policy* 14, no. 1 (December 1, 2014): 352–70. doi:10.1111/asap.12036.

¹³ "Facebook Users Worldwide 2016." *Statista*. Accessed October 8, 2016. <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>.

¹⁴ "Privacy Policy." *Twitter*. Accessed October 8, 2016. <https://twitter.com/privacy>.

addition, assuming a licensing agreement has been made for legal purposes, the widespread use and acknowledgment of these documents would lead to a successful strategy of informing authors of informed consent. When sites update their Terms and Conditions, many send an email blast that discusses what the changes are and how they affect the everyday use of the site. A similar notice could be sent out to inform users that their material will be archived.

Real World Examples

There are several case studies that carefully balance access and privacy, and these will be priceless when it comes to evaluating the practical boundaries of new media archiving.

The most directly related to social media archiving is the impending archive at the Library of Congress. In April 2010, Twitter gave all tweets to the Library of Congress (LoC) to be archived. After archiving all “old” tweets, LoC proceeded to advance onto more modern tweets, reaching 170 billion by the end of 2012¹⁵. With the goal to build a sustainable and chronologically organized archive, LoC aims to provide researchers with access to this archive to further scholarship and the distribution of information.

However, Data Scientist Dr. Katrin Weller travelled to LoC to hold a research fellowship on the Twitter archive from January to June 2015, but was unable to access any of the data. She notes “...I can say the archive isn’t going to happen ... even in the next 10 to 20 years ... I wouldn’t rely on this being available. The library is also

¹⁵ Library of Congress. “Update on the Twitter Archive At the Library of Congress,” January 1, 2013. http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf.

archiving text-based tweet formats ... you won't get the look and feel of what the platform looked like in ... and images and URLs inside a tweet may lose their value if they can't be resolved"¹⁶.

6 years later, and LoC is still unable to demonstrate a precedent for this type of archive, long after establishing that there is a need for an archive of this content. This lack of image and url preservation is a huge issue as many tweets are linked to other information. When using a tweet, you are able to upload your own content, or likewise attach media that Twitter offers, such as .gifs. The lack of these essential elements will render many tweets entirely useless. There are memes (an entertaining image or element of culture that is often changed slightly to reflect different circumstances, and is spread virally throughout the internet) that are left either lost or without context, and these are essential to our understanding of the internet and modern culture. In this case, the LoC as a formative archive has been paralyzed by the task of archiving only one facet of social media. Although it is impossible to tell whether ethics or law factored into their deliberation and paralysis, the circumstance nonetheless proves that there are certainly many more mountains to summit before creating a fully functional new media archive.

WITNESS, an international human rights organization that teaches untrained citizens of countries around the world how to document human rights violations. The intention of this organization is to shed light on violations that occur every day, and WITNESS contains an archival division in order to help preserve these videos, as well.

¹⁶ "Taming Twitter." *Research Information*, March 24, 2016.
<https://www.researchinformation.info/feature/taming-twitter>.

Due to the often controversial nature of the videos that they experience on a daily basis, WITNESS has a code of ethics that they maintain in order to protect both those recorded, and those who are behind the camera¹⁷. They provide examples for what circumstances would call for which type of data manipulation and obscurity, highlighting that the safety of all people involved is of the utmost importance.

In order to accomplish this, WITNESS utilises a suite of technological tools to intentionally obscure both data and metadata from the videos recorded. This includes an easily accessible app that ensures that all videos are secured and encrypted¹⁸, and another which blurs faces past the point of facial recognition software¹⁹. As an alternative and more conservative strategy, WITNESS comments that “You can choose not to show the footage, and instead provide your audience with a description of it”²⁰. They also recognize that it is not always possible to achieve informed consent, and for this they acknowledge the Professional Standards for Protection Work, which note that “When such consent cannot be realistically obtained, information allowing the identification of victims or witnesses should only be relayed in the public domain if the expected protection outcome clearly outweighs the risks”²¹.

¹⁷ “Video as Evidence: Ethical Guidelines.” *WITNESS*, September 22, 2015. <https://library.witness.org/product/video-as-evidence-ethical-guidelines/>.

¹⁸ Project, The Guardian. *CameraV: Secure Visual Proof* (version 0.2.6). Android 4.1 and up. The Guardian Project, 2015.

<https://play.google.com/store/apps/details?id=org.witness.informacam.app&hl=en>.

¹⁹ Project, The Guardian. *ObscuraCam* (version 2.0-RC2b). Android 2.3.3 and up. The Guardian Project, 2013. <https://play.google.com/store/apps/details?id=org.witness.sscphase1&hl=en>.

²⁰ “Video as Evidence: Ethical Guidelines.” *Library*, September 22, 2015.

<https://library.witness.org/product/video-as-evidence-ethical-guidelines/>.

²¹ “Professional Standards for Protection Work Carried out by Humanitarian and Human Rights Actors in Armed Conflict and Other Situations of Violence.” Accessed December 3, 2016.

Although WITNESS is video-focused and does not directly relate to social media archiving, they have set an excellent precedent in balancing the need for access with privacy and safety. Their recommended strategies are a video-parallel to my aforementioned suggestions regarding social media preservation (ethical obfuscation of identifying information versus metadata-only access).

A similarly parallel example of video-archiving is the use of police body cameras. Although there has been much discussion on the ethics of recording police encounters and the standards that all policemen and women should be held to regarding these recordings, the precedent that they are setting in regards to data storage and data policies are more relevant to this specific topic.

In our WITNESS example, most of their cases are small-scale, and specifically oriented towards appropriate measures to modify content to suit the privacy needs of few participants. However, when you consider body cameras, there is a lot more data involved that needs to be archived, and in this way, body camera data is more similar to predicaments one may encounter when dealing with social media content.

One of the unique and pertinent aspects of body cameras is that they require similar privacy modifications, but also absolutely necessitate access to workers and certain citizens, and therefore have very specific data requirements. This data includes tamper protections, a log of who has accessed the videos, and a strict authorization

matrix, indicating who can access the videos and for what time period²². These restrictions ensure security and privacy of the content, while maintaining good data storage practices.

This is a good strategy for a social media archive, as it displays a practical yet secure way to manage huge quantities of data. By controlling who has access they are maintaining a high degree of security for participants, but it is clear that as evidentiary materials, access must be an option as well.

The progress of body camera technology over the years also offers insight into the trials that storing large quantities of data entails. One example refers us to a blunder by the Oakland Police Department in 2014, where 25% of their body camera videos were mistakenly deleted in an attempt to update the system²³. This mistake was only revealed several years later, and naturally emphasizes the importance of securing data. Although this is clearly a huge issue, it's also important to realize that there are many equally bad, if not worse, outcomes for a similar scenario.

For example, instead of an accidental erasing, it's possible that all materials could have all their metadata stripped. Or worse yet, all materials would have privacy settings removed, instantly providing access to anyone who cared to view them.

Although I don't intend to be hyperbolic, this is a quintessential example of how it's vital

²² Community Oriented Policing Services, and Police Executive Research Forum. "Implementing a Body-Worn Camera Program," 2014.

<https://www.justice.gov/iso/opa/resources/472014912134715246869.pdf>.

²³ Veklerov, Kimberly. "Oakland Police Camera Gaffe: 25 Percent of Videos Deleted." *SFGate*. Accessed December 3, 2016.

<http://www.sfgate.com/crime/article/Oakland-police-camera-gaffe-25-percent-of-videos-9221030.php>.

to consider both the privacy and ethics implications in order to protect subjects of the archival material and their personal information.

This deletion mishap can be easily transferred and scaled to a social media archive. When considering worldwide social media penetration, Snapchat is on the low end with 200 million daily users as of 2016. Meanwhile, Facebook has 1.59 billion daily users²⁴. The concept of an archive accidentally erasing or publishing between 40 and 318 million user's personal information is terrifying, yet when you consider actual content generated, it becomes even more alarming. Every second, around 9000²⁵ snaps (the post-type of Snapchat) are sent. This means that every day, over 777 million individual items are produced for one singular social media platform. A quarter of this number would indicate that over 155 million of these posts could be (albeit only in the most distressing of situations) deleted, manipulated, or suffer from other varieties of data-loss. When dealing with these large numbers, it is important to consider the balance and compromise between privacy and access - archiving must be done when considering the privacy and safety of those involved, and access should only be given upon the assurance of this fact.

²⁴ "Global Social Media Statistics Summary 2016." *Smart Insights*, August 8, 2016. <http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>.

²⁵ Morrison, Kimberlee. "How People Are Really Using Snapchat (Infographic)." Accessed December 3, 2016. <http://adweek.it/1ONxt9V>.

Conclusion

Social media archiving is vital for our future, as our culture is leaning further into the use of social media as both a professional and amateur method of creating a cultural record. It is important to discuss the implications of archiving this information, considering the massive quantity of content being created on a daily basis.

After discussing different viewpoints and varying facets of social media archiving both legally and ethically, it is important to consider that the matter of privacy versus ethics are a balance, and an exercise in compromise. Although it may be legal in terms of copyright and privacy laws, it is imperative to consider the authors of the content, and those who would be affected by it. The ethics of preserving social media are dictated by the code that archivists follow, and as archivists, it is our responsibility to prevent any harm being incurred by our desire for access. Because of this, it is important to consider case studies and real world applications of this balance that have already negotiated these issues to form a suitable compromise between privacy and access.