

Ethical Challenges and Current Practices in Activist Social Media Archives

Ashlyn Velte

ABSTRACT

Social media (Web applications supporting communication between Internet users) empower current activist groups to create records of their activities. Recent digital collections, such as the digital archives of the Occupy Wall Street movement and the Documenting Ferguson Project, demonstrate archival interest in preserving and providing access to activist social media. Literature describing current practices exists for related topics such as Web and social media archives, privacy and access for digital materials, and activist archives. However, research on activist social media archives is scarce. These materials likely present subject- and format-specific challenges not yet identified in peer-reviewed research. Using a survey and semistructured interviews with archivists who collect activist social media, this study describes ethical challenges regarding acquisition and access. Specifically, the respondents were concerned about acquiring permission to collect and provide long-term access to activist groups' social media. When collecting social media as *data sets*, archivists currently intend to provide moderated access to the archives, whereas when dealing with social media *accounts*, archivists intend to seek permission to collect from the activist groups and provide access online. These current practices addressing ethical issues may serve as models for other institutions interested in collecting social media from activists. Understanding how to approach activist social media ethically decreases the risk that these important records of modern activism will be left out of the historical narrative.

© Ashlyn Velte.



KEY WORDS

Archival theory and principles, Ethics, Copyright and intellectual property, Privacy and confidentiality, Digital preservation, Social media archives, Web archives

Rapidly changing digital media present challenges for archivists intending to preserve them. Technological changes cause data corruption, media and software obsolescence, and inadequate metadata, which all lead to a “digital dark age,” a digital gap in the historical record.¹ Social media, as a form of digital material, also face these risks, particularly because they continuously develop. Also called Web 2.0 or social networks, social media are born-digital records of large-scale communication between communities and individuals on the Internet. Web 2.0 applications differ from Web 1.0 because they allow users to generate content and interact with one another rather than passively consuming existing content.² Social media applications serve a variety of purposes such as video sharing (YouTube and Vimeo), photo sharing (Flickr and Instagram), social networking (Facebook, LinkedIn, and Google+), blogging (WordPress and Tumblr), and microblogging (Twitter).³ All types of platforms involve liking, commenting, and sharing each other’s posts, making them highly interactive.

Because current events are increasingly documented on social media, preserving them has become a crucial part of archivists’ work. Recent activist movements use social media as platforms for citizen journalism; posts made on social media play an important role in organizing protests and sharing information about catalytic current events. Narratives about recent activist movements would differ without the social media record; social media allow events to be seen from the perspective of the protesters themselves, instead of relying on images and video captured by conventional media, thereby expanding the narrative consumed by the general public and consequently the narrative most often archived.⁴ In some cases, social media conversations bring underreported social issues into greater public consciousness, such as police violence against African Americans.⁵

Several online projects already collect social media from activist events or organizations. Collections of digital ephemera from the Occupy Wall Street movement include Emory University Libraries’ Archive of Occupy Wall Street tweets⁶ and a digital archives hosted by volunteers at George Mason University and the Roy Rosenzweig Center for History and New Media.⁷ Similar digital archives also document more recent social movements, such as the Documenting Ferguson Project at Washington University in St. Louis,⁸ the Baltimore Uprising Omeka site at the Maryland Historical Society,⁹ and A People’s Archive of Police Violence run by the Cleveland community and volunteer archivists.¹⁰

The last several years also saw activist groups form on college and university campuses. During the fall of 2015 alone, several campuses made national and international news when students held events highlighting the lack of diversity in higher education. Protests at Yale University, the University of Missouri, Ithaca College, and Claremont McKenna College resulted in real recognition and action taken by the college administrations.¹¹ Some of this campus activism

inspired archives at well-known universities (including Princeton University,¹² Harvard University,¹³ and UCLA¹⁴) to start collecting initiatives focused on student activism and its influence. In an effort to prevent a “digital dark age” and encourage others to begin documenting activism through social media, this article aims to describe current challenges and practices for activist social media archives. Surveys and semistructured interviews with archivists who collect activist social media reveal ethical challenges in acquisition and access, and identify current practices used to address those challenges.

Literature Review

WEB AND SOCIAL MEDIA ARCHIVES

Web archiving is the process of collecting and preserving portions of the World Wide Web in an archival format for future use.¹⁵ Several reports on Web archiving published over the last several years cover challenges and practices faced by Web archives.¹⁶ Most often Web archives use a Web crawler to capture URLs and display Web pages as they existed at the time they were captured. Collections with Web archives must choose from a plethora of tools with differing degrees of technical difficulty to harvest, preserve, and provide access to Web content.¹⁷ Most institutions manage their Web archives using external infrastructures like Archive-It, the most commonly used Web archiving service, which was created by the Internet Archive.¹⁸

Web archives vary widely in their permission-seeking practices. Reports on whether institutions seek permission to harvest Web sites from their creators show a variety of responses (from never to always),¹⁹ with few institutions implementing policies about when to seek permission.²⁰ Despite this, most Web archives have some policy for providing access to Web content. Most embargo access for a certain period of time (69%) while others employ no embargo (27%).²¹ However, when collecting social media, Web archives expressed greater ambiguity in both seeking permission from content creators and providing access.²²

Social media archives are one type of Web archives because social media are found online. Social media archives scope collections one of two ways: by limiting collections to specific user accounts or by hashtags or keywords on a theme.²³ Social media archives differ from Web archives in the opportunity they present to collect large data sets. This introduces challenges for collection development, access, and use that come with big data.²⁴ Unlike Web archiving more generally, collecting social media as data sets usually involves using the social media platform’s application programming interface (API) to harvest content and associated metadata in formats like JSON or XML.²⁵ Having access to social

media data sets provides researchers with the ability to analyze content and metadata on a large scale to determine community, national, or international trends.²⁶

Cultural heritage institutions collecting social media data sets face the challenge of changing terms of service for each platform, which makes it hard to maintain consistent collecting and access policies that are also legal.²⁷ Social media platforms are motivated to sell their data, which usually results in platforms' terms of service restricting how data are shared once they leave their system and limiting how often data can be called through their API. Another problem for archives is storage and access to large data sets.²⁸ It was widely publicized, for instance, that Twitter gave the Library of Congress all its public tweets from before 2010, but the tweets are not yet accessible due to difficulties indexing and searching vast amounts of data.²⁹

Social media collections also experience legal and ethical concerns beyond mercurial terms of service. North Carolina State University Libraries' Social Media Archiving Toolkit reviews these legal and ethical concerns. The toolkit explains that intellectual property law permits the preservation of copyrighted materials, while using the copyrighted materials in other creative works requires permission from the rights holder. Areas of ethical ambiguity include privacy, research use, and consent to collect. The toolkit concludes, "There is not a one-size-fits-all solution to solving the problems that present themselves when collecting social media."³⁰ Nonetheless, archivists balance privacy and access in collections regularly and can consult professional ethics³¹ and existing case studies to guide their decisions.³²

ACTIVIST ARCHIVES

In a literature review of archival approaches to social justice, Ricardo L. Punzalan and Michelle Caswell described increased efforts by archivists starting in the 1970s to include marginalized groups. The efforts have redefined archival concepts and training to include community archives and human rights archives. These changes have increased archival activism, or archival collections with social justice as a focus.³³ For Terry Cook, this represented a shift in the "archival paradigm" to a community archiving model³⁴ which has turned many traditional archiving institutions toward collecting materials from activists.³⁵ Activist archives share similarities with community archives³⁶: community groups create them, they are participatory in nature, and they intend to subvert dominant historical narratives.³⁷

Like some community archives, activist archives can be maintained by members of the activist community outside of existing archives to maintain autonomy over their records.³⁸ For example, the Lesbian Herstory Archive in

New York City is housed within lesbian community spaces for the purpose of defining lesbian history in contrast to the patriarchal narratives that dominate current conversations.³⁹ Activist archives intend to enact the social justice purpose of their communities.⁴⁰ In contrast, collections originating in activist communities also sometimes reside within existing libraries and archives. The Joseph A. Labadie Collection at the University of Michigan, for example, collects materials from labor movements, LGBTQ+ movements, antiwar movements, and student protests, among others.⁴¹ Also, the Tamiment Library and Robert F. Wagner Archives at New York University document labor movements and the American left.⁴² Though existing archives provide both stability for collections and visibility for the communities, and are mutually beneficial, documenting activists within existing structures may not always be positive for their communities. Some archives documenting indigenous groups confront issues of custody, colonialism, and displacement.⁴³ Other archives have been used by state agencies to conduct surveillance against vulnerable groups. One notable example is the Belfast Project's oral histories at Boston College, containing accounts from members of the Irish Republican Army (IRA), the Irish National Liberation Army (INLA), and other groups from both sides of the conflict in Northern Ireland. The oral histories were subpoenaed by the federal government on behalf of the Police Service of Northern Ireland despite Boston College implementing an embargo to protect documented individuals. Though the court limited the number of oral histories released under the subpoenas, the case demonstrates that archives may unintentionally implicate the people they document.⁴⁴ A similar case in Canada occurred when police archives containing records of activism were used to conduct surveillance on the LGBTQ+ community.⁴⁵ In response to instances like these, the archival profession is struggling to document sensitive groups without unintentionally endangering communities.

These challenges in activist archives raise concerns about archives' responsibilities to both privacy and access. One of the Society of American Archivists (SAA) Core Values is access and use, which involves promoting accessibility of materials to the largest audience possible.⁴⁶ The *SAA Code of Ethics* also includes access and use by calling for archivists to limit restrictions to materials as much as possible, while also protecting privacy for those documented in collections.⁴⁷ Balancing access and privacy is a fundamental exercise for archival professionals. Access policies take into consideration applicable privacy laws and relevant donor restrictions to help archivists make consistent equitable decisions in providing access to materials.⁴⁸ Balancing privacy and access becomes more complicated for digital records like social media because they may face ownership uncertainty and are sometimes already accessible over the public Internet.⁴⁹

If the challenges experienced by Web archives and activist archives are similar for activist social media archives, institutions may be deterred from

collecting activist social media without existing models. This study begins to fill this gap by describing ethical challenges and current practices experienced by pioneering archivists who collect activist social media.

Methodology

DATA COLLECTION

I modeled the surveys and interviews in this exploratory study on the survey and semistructured interview guide developed by Lisl Zach and Marcia Frank Peri.⁵⁰ Like this study, they identified challenges and practices for electronic records management, a new area in archives at the time. They also used the same data collection methodology as this study did. The questions in Gail Truman's report on Web archiving helped guide the questions during the semistructured interviews because its subject was similar to the current study.⁵¹

The participants first took a ten-to-twenty-minute online survey sent out through professional networks and listservs. The link was active for one month. After respondents consented to participate, they answered both multiple choice (some allowing for multiple responses) and open-ended survey questions (see Appendix A). Though thirteen participants completed the survey, they did not answer every question. After completing the survey, I asked participants if they wanted to participate further in a thirty-to-sixty-minute semistructured interview via telephone. Questions during the interview covered project workflows, challenges, and how they approached those challenges (see Appendix B).

To preserve anonymity, I did not ask specific questions about archivists' employing institutions during data collection. Few libraries and archives collect activist social media content, so it might have been possible to identify institutions or individuals from their responses. Based on the qualitative responses on the survey, universities employed at least some of the participants. One respondent stated, "We are not a library," implying that that respondent may work at a museum or other similar cultural heritage institution. All three participants who agreed to semistructured interviews happened to be digital archivists at universities.

DATA ANALYSIS

I used descriptive statistics for each quantitative question on the survey. Analysis of open-ended survey and interview questions followed Yan Zhang and Barbara M. Wildemuth's⁵² suggested practices, which include the following steps: prepare the data, define the unit of analysis, develop categories and a coding scheme, test coding scheme, code all the text, assess coding consistency, draw

conclusions from the coded data, and report methods and findings. I recorded and transcribed interviews with consent from the participants. I developed and applied the coding scheme based on common topics that developed across participants' data. The initial coding scheme had six variables (collection development and acquisition, access, project management, acquiring permission or consent, challenges, outcomes, and areas of future research). The codes were further subgrouped into two broader categories that developed across survey and interview responses (ethical challenges and current practices). Direct quotes reported in the results reflect the most concise and representative portions of the interviews. I chose them carefully to avoid identifying interview participants.

Results and Discussion

The surveys and interviews identified similar challenges and practices across participating activist social media archives. The responses described both ethical challenges faced during acquisition and access and current practices that address those challenges.

To contextualize the interview results reported in this study, the three interview participants described their activist social media collections. Archivist 1 collects social media at a library and has been doing so for several years. A violent event affecting the campus and local community catalyzed Archivist 1 to harvest Twitter and Instagram data from a specific hashtag related to this event. Because of the hatred involved with the event, many posts were related to activist or advocacy messages documenting both sides of the social conflict. The collecting project related to the event did not start as a specific activist-oriented collecting initiative and is not focused on particular activist groups.

Archivist 2 witnessed increased activism on the university's campus during 2014 and 2015. Student groups held sit-ins on campus in the fall of 2015. Archivist 2 launched an initiative to collect records from activist student organizations for the university archives, holding two collection drives, one in the student center and one in the archives, which is farther away from the center of campus and offered some anonymity. When the archives advertised these drives, it emphasized that it would accept any kind of record, from traditional analog materials to email archives and social media accounts, which resulted in the acquisition of several hybrid collections from student groups.

Archivist 3 also observed an increase in activist activity during fall 2015. Archivist 3 was concerned about losing valuable information about these events, so Archivist 3 began to collect student papers including social media content. At the time of the interview, Archivist 3 had only recently begun publicizing this student-activist collecting initiative and had not yet received new materials though intended to collect social media as part of this initiative.

ETHICAL CHALLENGES

One might expect that activist social media collections would report technological difficulties when acquiring social media data. However, interview and survey responses implied that data collection is not the most challenging aspect for activist social media archives. Rather, they face legal and ethical challenges regarding acquisition strategies for and providing access to social media content.

When asked to indicate challenges experienced from a list (see Figure 1), most survey participants reported legal issues (5 respondents; 63%) and ethical issues (4 respondents; 50%). Similar issues were brought up by those experiencing “other” challenges (3 respondents; 38%). When these three respondents were asked to explain what other challenges they experienced, two of them described their issues as “curatorial,” implying issues choosing what content to collect. The rest of the participants cited issues outside of acquisition concerns: three reported harvesting data as a challenge (38%); and two found funding their projects difficult (25%).

Survey respondents also identified ethical concerns when asked about areas of future research in activist social media archiving. As Table 1 indicates, only one participant was curious about the usability and value of different harvesting tools. The most-mentioned area of future study was ethical practices when collecting social media. They specifically named ethical issues surrounding consent and privacy for the activists whose social media they collect. Two participants called for creating best practices surrounding this issue.

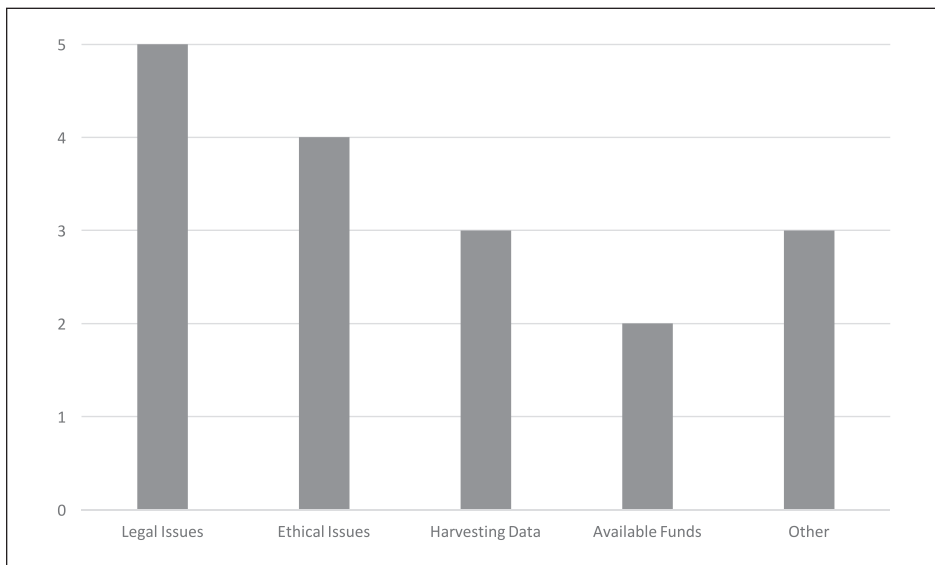


FIGURE 1. This graph illustrates the number of participants who experienced different challenges.

Table 1. Areas of Future Research Identified by Participants

"Mainly of the tools are evolving in use and value."
"I myself am curious about what the use cases in the future for Facebook pages will be. Is it going to be drilling down and seeing the exchanges of annoyance and anger about social issue X on the Facebook of some organization? Or use of a large set of Facebook pages for text analysis? Or evidentiary, that is, to demonstrate that despite statements to the contrary, organization X did say Y on date Z. For a particular organization, I can see that having an accurate archive of the organization's own Facebook etc. can be useful if not critical for evidentiary and records management type reasons. However if we are talking about large publishing organizations, I am not so sure (as I have already suggested) that worrying about social media isn't a distant second to the need to archive the organization's site fully and accurately, which I think is assumed to be simpler than it often is for organizations with big sites."
"Data preservation and privacy issues are always tough."
"More attention needs to be invested into obtaining consent of content creators."
"We'd like to perform a more thorough analysis of the ethical considerations of harvesting and preserving social media data, especially of social media users who are not aware of our collecting initiative. Generally speaking, the archival community could develop some 'better practices' when it comes to collecting, describing, and providing access to these sorts of data. How institutions deal with copyright and privacy for harvested social media varies quite a bit. I'd like to see some more defined best practices for this."

During the interviews, archivists were most eager to discuss ethical issues, often returning to the topic at various points in the conversation. Archivist 1 identified broad ethical issues experienced when collecting social media saying, "Getting the stuff is relatively easy. But all of the rules and policies and protocols that support it, those are a little bit trickier." Those rules and policies are the professional ethics and local policies that balance privacy and access. Archivist 1 elaborated, saying, "[Harvesting social media] is a fairly conclusive fair use argument. It's a pretty legal thing to do, I think that there will be some issues about how to interpret certain terms of service. But doing it ethically [is less clear]." It is telling that despite actively collecting and preserving data sets of activist social media, Archivist 1 had yet to open the data for public access. Archivist 1 said, "Because of rights issues, and because of terms of service, we're very much likely to provide access in a controlled manner." In the past, this archives provided opt-out for Instagram users. However, Archivist 1 said:

If we were to do that with collections that had more sensitive information, like an activist or advocacy oriented collection we might lose a lot of data. Because people would write and say "I don't want it to be part of your archive." . . . It's something we're chewing over because it gets into professional ethics.

For that reason, Archivist 1 was carefully considering alternative models of access for the data collected, rather than making the data widely available online.

Archivists 2 and 3 indicated that they tailored acquisition practices to permit online access. When asked about challenges they faced, Archivist 2 expressed concerns about making student groups vulnerable to surveillance: "We were aware of the power dynamics at play and we're aware that we're

dealing with undergraduate students who on top of leading different initiatives through their organization they're full time students." Archivist 2 was also concerned that collecting activist social media while events are happening "is creating more silences, creating more gaps, creating more vulnerabilities for people to be surveilled, to be harassed." Archivist 2 believed it could be more beneficial to develop long-term partnerships with groups before their activities become "trendy" to avoid further harm.

Similar to Archivist 2, Archivist 3 was concerned about privacy saying, "We're definitely going to be seeking consent, but we were just concerned about preservation first and access later. . . . That was a concern about privacy, and if they say no we're going to delete it anyway." Archivist 3 sought permission to ensure that the activist student groups' intentions were to provide long-term access to their social media records. Archivist 3 explained that the institution should always have been collecting student records from activist groups, and that starting its specific initiative was one way to diversify its collections, to fill in existing gaps in the collections. Archivist 3 intends to approach the student groups more actively, instead of passively waiting for people to deposit their materials. "Outreach really is a big part of building the relationships and making sure that they feel comfortable depositing their records with us." Archivist 3 stressed the importance of building relationships with activists so they will trust that the archives' goal is to preserve the historical record and not to perform surveillance for the university.

These challenges regarding ethical acquisition and access fit what is known about social media archives. For Archivist 1, the problem was the scalability of permission-seeking with large data sets of social media content. Research on social media archiving indicates that data analysis might reveal personal information about individuals who may not know how their information is being used. Nevertheless, it remains unclear if collecting institutions have an ethical responsibility to seek consent for data sets where privacy issues may arise.⁵³

The fact that all interviewed archivists expressed concern about professional ethics reflects their understanding of activist archives. Since relationships between activist groups and collecting institutions risk maintaining oppressive social precedents and introducing opportunities for surveillance,⁵⁴ it seems natural that archivists familiar with these issues might want to build relationships with activist groups. According to the interviewed archivists, permission to collect helps mitigate causing further harm.

CURRENT PRACTICES

Interviewed archivists' chosen acquisition strategies informed both permission-seeking and access practices. As described earlier, Archivist 1 scoped

their social media collections using hashtags. Research on social media archives reports that other archives also scope their social media collections using hashtags, keywords, or geolocations. Another strategy is to limit the scope to particular accounts.⁵⁵ While both strategies may result in content loss, without planning, social media collections can capture too much or too little.⁵⁶ Archivist 1 said that they “align decisions about what data to harvest, with our established collecting strengths.” Limiting acquisitions only to particular hashtags on Instagram and Twitter ensured that Archivist 1 only preserved data sets relevant to the institution’s collection development policy.

Archivists 2 and 3 chose the other method of scoping social media collections: they acquire social media accounts rather than hashtags. Because of a concern with surveillance, Archivist 2 took particular care to ensure anonymity for students in activist groups. For instance, Archivist 2 hosted two collection drives related to the archives’ activist student group collection initiative: one in the highly visible student center and another in a less public space (the archives building). When collecting social media from activist groups, Archivist 2 only collected accounts and not related hashtags. Archivist 2 made this decision because of intentions to acquire permissions from activist groups to collect, preserve, and provide long-term access to their social media content. Archivist 2’s university collected social media using Archive-It, which provides access to Web archives using the Wayback Machine. Because anyone can access the Wayback Machine, obtaining permissions from activist groups was important to Archivist 2, who explained the choice to follow accounts instead of hashtags:

We didn’t want a situation where we had a number of tweets that were not intended necessarily to end up in the archives. So we decided to go down a traditional provenance based approach and directly inquire to student organizations about permission to capture their site and make it available. And all of them agreed.

Similarly, Archivist 3 intended to acquire permission to make activist student groups’ materials available to the public: “We will have a form they will sign. And what we’ve decided is that we will ask the leader of each student organization that’s responsible for the [social] media that we’ve captured to discuss among their group and then sign off on behalf of the entire group.” Like Archive-It, the tool Archivist 3 used to capture social media (Perma.cc) also makes preserved materials widely available over the Internet.

Survey responses in this study indicated that most participants are likely to collect social media Web pages rather than datasets. Archive-It was the most commonly used tool among the participants for capturing social media (see Figure 2). Of the eight archivists who answered the question on tools, four responded that they use Archive-It (50%), and two use personal archives that have been downloaded and donated (25%). The other tools—Twarcc (13%), Lentil

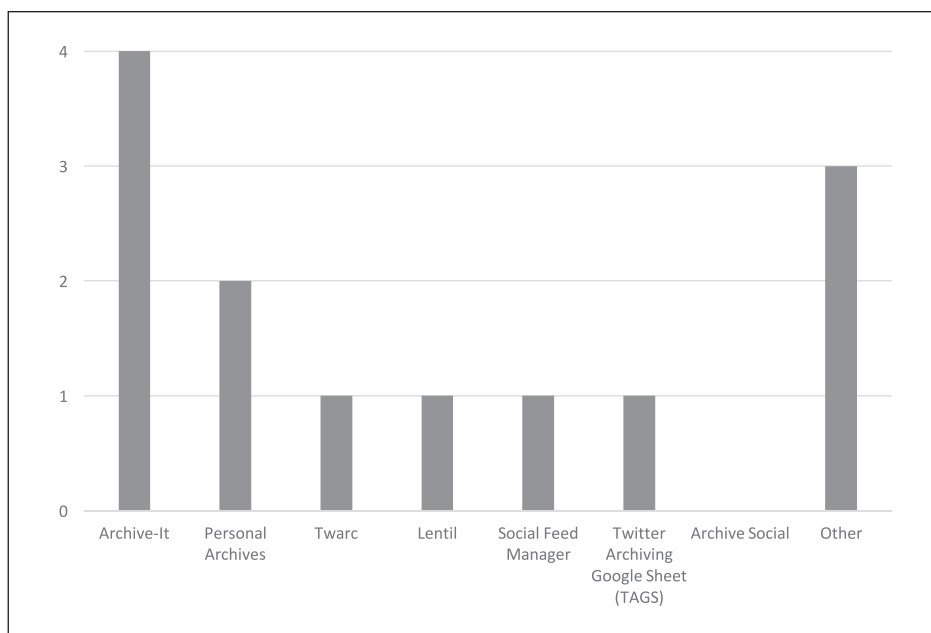


FIGURE 2. This chart illustrates the number of participants who used different social media harvesting tools.

(13%), Social Feed Manager (13%), and Twitter Archiving Google Sheet (TAGS; 13%)—collect data sets and not Web pages. No respondents used ArchiveSocial (0%). The second most frequently reported tool used was “other” (38%). When asked to explain, one archivist (13%) said that they “Use the Internet Archive as a contractor for Web archiving, which is much like Archive-It but not the same thing.” The other respondents that chose “other” reported using Perma.cc (25%), which also captures Web pages.

This replicates findings from Web archiving reports indicating that a large number of Web archives rely on Archive-It for collection, preservation, and access.⁵⁷ The results of this survey demonstrate that Archive-It and tools that capture social media as Web pages (i.e., personal archives and Perma.cc) are used more often for collections of activist social media than tools that collect data sets (e.g. Lentil, Social Feed Manager, Twarc). This implies that most activist social media collections more likely share similarities with the acquisition and access strategies employed by Archivists 2 and 3.

Archivists 2 and 3, who both collected accounts, intended to make activist social media available as archived Web pages using the publicly available access layer provided by their acquisition tools (Archive-It and Perma.cc respectively). However, for archivists collecting social media data sets, it is more difficult to acquire permission to collect from every social media user. To make data sets accessible, Archivist 1 anticipated providing moderated access by requiring

researchers to view data either in the reading room or by providing tweet IDs that the researcher could then “hydrate” with the tweet content through the API. Hydration is one way that researchers and archivists provide ethical access to social media content.⁵⁸ This method allows archives to preserve a list of tweet IDs that the researcher then populates (or “hydrates”) with the content of the tweet using Twitter’s API. Any tweets deleted since their IDs were archived have no content to hydrate the ID. Therefore, hydration reflects users’ choices to delete content after the tweet ID was archived. This comes with its own issues because deleted posts will not be preserved. However, it reflects Twitter users’ intentions as it only provides access to tweets currently available online, and it does not require archivists to contact every user in the data sets. This method of data sharing is also permitted under Twitter’s current terms of service. Similarly, for Instagram data, Archivist 1 said, “I don’t think we would make the Instagram content available on the open Web except for cases where we got permission from the people who either took or posted the photos.”

Models for moderated access exist among both Web archives and more general electronic records archives. Truman’s Web archiving report found that 36 percent of the twenty-three institutions surveyed provide access to Web archives in the reading room because of copyright or privacy concerns.⁵⁹ Some institutions also provide moderated access to electronic records. For example, the Salmon Rushdie Papers’ born-digital materials at Emory University are only available for research on a computer terminal in the reading room,⁶⁰ and the University of California, Irvine, developed a “virtual reading room” requiring registration that includes a terms of use to log-in and view digital materials.⁶¹

Both access models described by the interviewed archivists were developed to balance privacy and access, and replicate conditions of access for physical materials. Menzi Behrnd-Klodt described strategies for handling particular types of records to maintain this balance for digital materials. One strategy is to work with donors to identify privacy concerns.⁶² This approach is similar to Archivist 2’s and Archivist 3’s decisions to seek permission to provide access to materials from activist groups. However, for archivists like Archivist 1 who collect data sets, Behndt-Klodt noted that collecting social media is challenging because, although users post publicly, they do not always understand that their content is available for long-term preservation and use.⁶³ Understanding privacy laws, ethical codes, and institutional tolerance for risk allows archives to “develop a reasonable, thoughtful access policy that balances access and privacy.”⁶⁴ Archivist 1’s plan to allow moderated access to the social media archives reflects this thoughtful understanding of the issues present in archiving social media data sets and follows the existing models for access to digital collections described above.

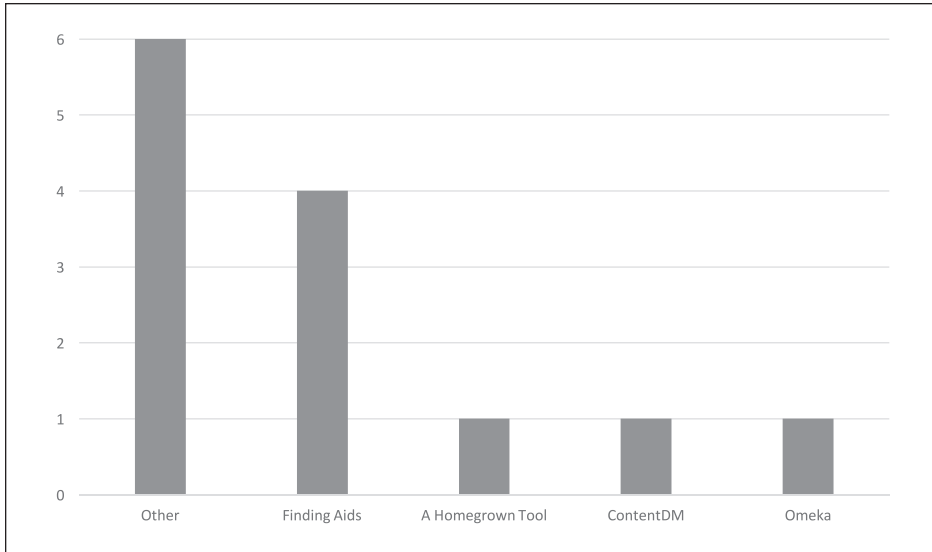


FIGURE 3. This chart illustrates the number of participants who used access tools.

Though the survey results do not indicate why decisions are made, they do indicate that activist social media archives choose different access methods. When asked about which access tools they used for their activist social media collections (see Figure 3), most archivists reported using “other” tools not named on the survey (6 of 8 archivists; 75%), likely because access layers provided by social media archiving tools like Archive-It were not among the survey options. Of the options provided, the most common way archivists provided access to activist social media is through archival finding aids. Four of the eight archivists who answered the question reported using finding aids (50%). Usefully, all four of them provided more information in the free response section about how researchers will view social media. One respondent has not begun providing access to social media content, but said that it will be discoverable in a finding aid and access will occur on-site in a reading room like the moderated access model described by Archivist 1. Another respondent revealed that they do not have a way of providing access to born-digital materials through the finding aid, explaining that “we provide access to digital images and some documents through finding aids, but we do not have a method for serving up born-digital material in any meaningful way yet (especially not social media content).” Two respondents said they have another discoverability tool in addition to finding aids; one uses the Wayback Machine and the other uses the public link provided by Perma.cc. In summary, half of the participating archivists described activist social media collections in archival finding aids, and one-quarter reported using online tools to display activist social media collections online.

The archivists interviewed in this study also developed strategies for overcoming their concerns about creating or filling gaps in the historical record. At the time of the interview, Archivist 3 intended to approach student groups instead of passively waiting for people to deposit their materials. Archivist 3 wanted to build collections by making it a “participatory conversation” between the archives and activist student groups and planned to host events with the groups on “a broader range of issues that relate to some of their activism.” One of these events would specifically discuss gaps in the archival record. Archivist 3 saw it as an opportunity to help activist groups understand that “whatever we do decide to keep . . . that’s a political act because it’s a level of interpretation. . . . I want to talk about the challenges to building a diverse collection.” For Archivist 3, these events between library professionals and student groups would hopefully “be part of building the relationship with students and making sure they feel comfortable depositing their records with us.”

Despite developing practices to address ethical challenges, all three archivists reported a desire for better professional guidance through best practices for collecting and providing access to activist social media. When asked about areas of future research, Archivist 1 said,

We’re not an insensitive group of people who are going to disregard ethics, but I think that there’s a pretty good momentum in favor of there being a way for it to be done ethically. We don’t have all of the answers of what that’s going to look like because in a way it’s a discussion that’s happening with today’s researchers who are accessing and using Twitter data.

Archivist 1 had the impression that the archival profession is interested in finding ways to ethically support activist social media archives. Similarly, Archivist 2 thought that having professional case studies that implement ethical collecting methodologies would be useful.

Indeed, other archivists are already developing practices for working with activist social media ethically. For example, Documenting the Now, a collaborative project between the University of Maryland, the University of California, Riverside, and Washington University in St. Louis funded by the Andrew W. Mellon Foundation, investigates the ethical collection, preservation, and access of social media data sets. Documenting the Now involves a variety of stakeholders, including researchers, to develop tools and best practices that reduce the ethical ambiguities described in this study.⁶⁵

Limitations and Areas for Future Research

The primary limitation of this study is its lack of generalizability. Because there were not enough participants to achieve statistical significance, one cannot

make assumptions about the results' representativeness. However, few archives were collecting activist social media at the time of data collection (early 2016). Taking this into consideration, having thirteen survey respondents was acceptable. Nonetheless, the study's purpose was to describe current approaches to thinking about ethical challenges. As Michelle Caswell, Marika Cifor, and Mario Ramirez explained, "Using semi-structured qualitative interviews is a well-established method in archival studies. . . . The resulting data are descriptive in nature; the goal of such research is to generate a 'thick description' of a particular phenomenon in a single setting."⁶⁶ This study makes no claims about the generalizability of its findings; rather, it describes current approaches to new professional problems.

Another limitation is the inherent subjectivity of coding qualitative data. Though I attempted to reduce subjectivity by coding more than once to account for changes in how the coding scheme was applied across participants, some subjectivity was still present. A final limitation is the exploratory nature of this study that described the existing landscape but did not evaluate the identified approaches.

Based on the limitations and the findings presented in the study, future research might focus on determining which characteristics of activist social media collections lead to the adoption of particular practices; statistically determining the most commonly adopted practices by including more participants, especially since more archives collect activist social media now; developing case studies of ethical practices in collecting activist social media; and assessing the benefits and drawbacks of particular access models described in this study.

Conclusion

This article describes challenges and practices for collections of activist social media. Specifically, the greatest challenges for the archivists in this study lie in ethically acquiring and providing long-term public access to activist social media archives. The participating archivists tend to make social media collections widely available online only after activist groups have knowingly consented to archiving their social media, while institutions collecting large data sets intend to provide moderated access because of terms of service and ethical questions about consent. In the absence of professional consensus, these thoughtful approaches that balance privacy and access serve as models but cannot be considered best practices.

As Archivist 1 explained and the Documenting the Now project demonstrates, determining ethical access and use will involve not only archivists but researchers as well. Continuing to work toward a better professional understanding of the complex ethical issues present in activist social media archives

will encourage other institutions to document activism. Better professional support decreases the risk that these important records of current events will be left out of the historical narrative. In the meantime, this article provides some examples of how archives currently approach collecting activist social media, which hopefully encourages better documentation of activism through social media.

Appendix A: Archivist Survey

1. Were there already existing digital archival collections at your library?

- ☐ Yes
- ☐ No
- ☐ Don't know

If so, what was it?

2. How was your social media collecting project funded?

- ☐ Grant
- ☐ It was free
- ☐ Money was allocated to it from the library or department budget
- ☐ Other

If other, please specify:

3. What social media platforms do you collect for your project?

- | | | |
|-----------------------------------|------------------------------------|-----------------------------------|
| <input type="checkbox"/> Facebook | <input type="checkbox"/> Vimeo | <input type="checkbox"/> Flickr |
| <input type="checkbox"/> Twitter | <input type="checkbox"/> Instagram | <input type="checkbox"/> LinkedIn |
| <input type="checkbox"/> Tumblr | <input type="checkbox"/> YouTube | <input type="checkbox"/> Blogs |
| <input type="checkbox"/> Vine | <input type="checkbox"/> Pinterest | <input type="checkbox"/> Other |

If other please specify:

4. What tools did you use to harvest social media?

- ☐ Twarc
- ☐ Lentil
- ☐ Feed Manager
- ☐ ArchiveSocial
- ☐ Archive-It
- ☐ Archiving Google Sheet (TAGS)
- ☐ Archives (Downloaded by individual users and donated to your collection)
- ☐ Other

If other, please specify:

5. What were some challenges you faced regarding platform specific formats, or the tools used to collect data? Check any that apply:

- ☐ Legal issues (such as intellectual property, and privacy questions)
- ☐ Ethical issues (such as privacy, personal information, and intentions of the social media user)
- ☐ Harvesting data (e.g. tools were too difficult to use)
- ☐ Available funds

6. What metadata do you use to help make your social media data more accessible? If you use a metadata standard what standard do you use? Check any that apply:

- ☐ DublinCore
- ☐ PREMIS
- ☐ Library Congress Subject Headings
- ☐ Folksonomies (such as user generated tags)
- ☐ MARC
- ☐ METS
- ☐ Other

If other, please specify:

7. What tools do you use to provide access to content? Check all that apply:

- ☐ Omeka
- ☐ ContentDM
- ☐ WordPress
- ☐ Homegrown tool
- ☐ Finding aids
- ☐ Other

If other please specify:

8. Name some of the outcomes you have experienced as a result of this project. Check all that apply:

- ☐ Wider use of analog and digital material
- ☐ Increased traffic to website or access platform
- ☐ Media attention
- ☐ Harassment
- ☐ Use in student projects
- ☐ Increased donations of material
- ☐ Increased monetary donations
- ☐ Increased attendance at library programs related to your social media collections
- ☐ Other

If other, please specify:

9. Is there any aspect of social media collecting that should be examined further by professionals and the professional literature?
10. Are you interested in being interviewed about your social media project?
- ☐ Yes
 - ☐ No

Appendix B: Semistructured Interview Guide

I. Work flows

- a. Why did you begin this project?
- b. How do collect social media content?
- c. How do you decide what to collect?
- d. What was important to you when making decisions about this project?
- e. How do you provide researchers with access? Can you describe the tools or process you use?
- f. Are there any issues you have had with providing patrons with access? If so have you created any policies to deal with these issues? What are they?
- g. On a fairly high level could you step me through the process of collecting social media for this process from start to finish?
- h. What were the greatest challenges you faced and how did you overcome them?
- i. What would have made these challenges easier for you?

II. Administrative concerns

- a. What were the campus-wide policies that supported the creation of this project?
- b. Is there dedicated funding? How did you find funding?
- c. What other staff is involved with helping run the project if any?
- d. What is the future direction of this project or related projects?
- e. Describe one outcome of the project? Do you have a particular story you'd like to share about an outcome?

III. Cooperation and coordination

- a. What stakeholders did you work with most closely on this project?
- b. What is your relationship with Library/Archives IT? How did they help you?
- c. What contact did you have with any legal team or institutional attorneys?
- d. Describe any legal issues you worked on if any.
- e. What is your relationship with the activist group involved? How did they help you if at all?

IV. Best practices

- a. What best practices would you like to see implemented surrounding the collection and preservation of social media?
- b. Is there anything else you would like to share about collecting social media or working with activist groups?

NOTES

- ¹ Stuart Jeffrey, "A New Digital Dark Age? Collaborative Web Tools, Social Media, and Long-Term Preservation," *World Archaeology* 44 (2012): 553–57, <http://doi.org/10.1080/00438243.2012.737579>.
- ² Daniel Zeng et al., "Social Media Analytics and Intelligence," *IEEE Intelligent Systems* 25, no. 6 (2010) 13–16, doi:10.1109/MIS.2010.151.
- ³ Zeng et al., "Social Media Analytics and Intelligence"; Sara Day Thomson, *Preserving Social Media*, DPC Technology Watch Report (Great Britain: Digital Preservation Coalition, 2016), <http://dx.doi.org/10.7207/twr16-01>.
- ⁴ Bergis Jules, "Hashtags of Ferguson," *Medium*, <https://medium.com/on-archiv/hashtags-of-ferguson-8f52a0aced87>.
- ⁵ Deen Freelon, Charlton D. McIlwain, and Meredith D. Clark, *Beyond the Hashtags: #Ferguson, #Blacklivesmatter, and the Online Struggle for Offline Justice* (Washington, D.C.: Center for Media and Social Impact, 2016), http://cmsimpact.org/wp-content/uploads/2016/03/beyond_the_hashtags_2016.pdf.
- ⁶ Leslie King, "Emory Digital Scholars Archive Occupy Wall Street Tweets," *Emory Report*, September 21, 2012, http://news.emory.edu/stories/2012/09/er_occupy_wall_street_tweets_archive/campus.html.
- ⁷ John Erde, "Constructing Archives of the Occupy Movement," *Archives and Records* 35 (2014): 77–92, <http://doi.org/10.1080/23257962.2014.943168>.
- ⁸ LaTanya Buck et al., *Documenting Ferguson: Project Explanation and Purpose*, Documenting Ferguson Committee (2014), <http://digital.wustl.edu/ferguson/DFP-Plan.pdf>.
- ⁹ Maryland Historical Society, "Announcing BaltimoreUprising2015.org," <http://www.mdhs.org/announcing-baltimoreuprising2015org>.
- ¹⁰ A People's Archive of Police Violence in Cleveland, "Support Us," <http://www.archivingpoliceviolence.org/support>.
- ¹¹ Taylor Maycon, "Ithaca College President to Resign Following Student, Faculty Backlash," *USAToday College*, January 14, 2016, <http://college.usatoday.com/2016/01/14/ithaca-college-president-resigns/>; David Smith and Steven W. Thrasher, "Student Activists Nationwide Challenge Campus Racism—and Get Results," *The Guardian*, November 13, 2015, <http://www.theguardian.com/us-news/2015/nov/13/student-activism-university-of-missouri-racism-universities-colleges>.
- ¹² Jarrett Drake, "Announcing ASAP: Archiving Student Activism at Princeton," *Mudd Manuscript Library Blog*, December 2, 2016, <https://blogs.princeton.edu/mudd/2015/12/announcing-asap-archiving-student-activism-at-princeton/>.
- ¹³ Jessica Farrell, "Archiving Student Action at HSL," *Et Seq. The Harvard Law School Library Blog*, February 11, 2016, <http://etseq.law.harvard.edu/2016/02/archiving-student-action-at-hsl/>.
- ¹⁴ Kartik Kolachina, "Student Groups Preserve Their History with UCLA's New Archive Project," *Daily Bruin*, April 10, 2015, <http://dailybruin.com/2015/04/10/student-groups-preserve-their-history-with-ucla-new-archive-project>.
- ¹⁵ "Web Archiving," International Internet Preservation Consortium (2017), <http://netpreserve.org/web-archiving/>.
- ¹⁶ Gail Truman, "Web Archiving Environmental Scan," *Harvard Library Report* (2016), <https://dash.harvard.edu/handle/1/25658314>; Jefferson Bailey et al.; *Web Archiving in the United States: A 2013 Survey* (NDSA Report, 2014), http://www.digitalpreservation.gov/documents/NDSA_USWebArchivingSurvey_2013.pdf; National Digital Stewardship Alliance (NDSA), *Web Archiving Survey Report*, 2012, http://www.digitalpreservation.gov/documents/ndsa_web_archiving_survey_report_2012.pdf.
- ¹⁷ Truman, "Web Archiving Environmental Scan"; Bailey et al., *Web Archiving*.
- ¹⁸ NDSA, *Web Archiving Survey Report*.
- ¹⁹ NDSA, *Web Archiving Survey Report*, 6; Bailey et al., *Web Archiving*, 12–13.
- ²⁰ NDSA, *Web Archiving Survey Report*, 6–7.
- ²¹ Bailey et al., *Web Archiving*, 15.
- ²² Bailey et al., *Web Archiving*, 22.

- ²³ Thomson, *Preserving Social Media*, 22–23.
- ²⁴ Thomson, *Preserving Social Media*, 4.
- ²⁵ Thomson, *Preserving Social Media*, 7.
- ²⁶ North Carolina State University Libraries, Social Media Archiving Toolkit, “Social Media Data Research and Use” (2015), <https://www.lib.ncsu.edu/social-media-archives-toolkit/research-and-use/research>; Thomson, *Preserving Social Media*.
- ²⁷ Thomson, *Preserving Social Media*, 15–16.
- ²⁸ Thomson, *Preserving Social Media*.
- ²⁹ Library of Congress, “Update on the Twitter Archive at the Library of Congress,” white paper, 2013, https://www.loc.gov/static/managed-content/uploads/sites/6/2017/02/twitter_report_2013jan.pdf.
- ³⁰ North Carolina State University Libraries, Social Media Archives Toolkit, “Legal and Ethical Implications,” 2015, <https://www.lib.ncsu.edu/social-media-archives-toolkit/legal>.
- ³¹ Society of American Archivists, *SAA Core Values Statement and Code of Ethics*, 2011, <http://archivists.org/statements/saa-core-values-statement-and-code-of-ethics>.
- ³² For example, *Case Studies in Archival Ethics* (Chicago: Society of American Archivists: 2017), <https://www2.archivists.org/groups/committee-on-ethics-and-professional-conduct/case-studies-in-archival-ethics>.
- ³³ Ricardo L. Punzalan and Michelle Caswell, “Critical Directions for Archival Approaches to Social Justice,” *Library Quarterly* 86 (2016), doi:10.1086/684145.
- ³⁴ Terry Cook, “Evidence, Memory, Identity, and Community: Four Shifting Archival Paradigms,” *Archival Science* 13 (2013): 95–120, doi:10.1007/s10502-012-9180-7.
- ³⁵ Alycia Sellie et al., “Interference Archives: A Free Space for Social Movement Culture,” *Archival Science* 15 (2015): 453–72, doi:10.1007/s10502-015-9245-5.
- ³⁶ Sellie et al., “Interference Archives,” 457.
- ³⁷ Andrew Flinn and Mary Stevens, “‘It is no mistri, wi mekin histri.’ Telling Our Own Story: Independent and Community Archives in the UK, Challenging and Subverting The Mainstream,” *Community Archives: Shaping of Memory* (London: Facet Publishing: 2009), 3–27; Michelle Caswell, “Toward a Survivor-centered Approach to Records Documenting Human Rights Abuse: Lessons from Community Archives,” *Archival Science* 14 (2014): 307–22, doi:10.1007/s10502-014-9220-6.
- ³⁸ Shauna Moore and Margaret Pell, “Autonomous Archives,” *International Journal of Heritage Studies* 16 (2010): 255–68, <http://dx.doi.org/10.1080/13527251003775513>.
- ³⁹ Joan Nestle, “The Will to Remember: The Lesbian Herstory Archives of New York,” *Feminist Review* 34 (1990): 86–94, doi:10.2307/1395308.
- ⁴⁰ Sellie et al., “Interference Archives,” 457.
- ⁴¹ Julie A. Herrada, “Joseph A. Labadie Collection,” University of Michigan Library, <https://www.lib.umich.edu/labadie-collection>.
- ⁴² New York University, “Tamiment Library and Robert F. Wagner Archives: History and Description,” <https://www.nyu.edu/library/bobst/research/tam/history.html>.
- ⁴³ Evelyn Wareham, “Our Own Identity, Our Own Taonga: Voices in New Zealand Record-keeping,” *Archivaria* 52 (2001): 26–46; Jeannette Allis Bastian, “A Question of Custody: The Colonial Archives of the United States Virgin Islands,” *The American Archivist* 64, no. 1 (2001): 96–114.
- ⁴⁴ James Allison King, “‘Say Nothing’: Silenced Records and the Boston College Subpoenas,” *Archives & Records* 35 (2014): 28–42, <http://dx.doi.org/10.1080/23257962.2013.859573>; Krista White, “Minding the Gaps: Interprofessional Communication and the Stewardship of Oral Histories with Sensitive Information,” *Journal of Academic Librarianship* (2017), <https://doi.org/10.1016/j.acalib.2017.06.007>.
- ⁴⁵ Steven Maynard, “Police/Archives,” *Archivaria* 68 (2009): 159–82.
- ⁴⁶ Society of American Archivists, *SAA Core Values Statement and Code of Ethics*.
- ⁴⁷ Society of American Archivists, *SAA Core Values Statement and Code of Ethics*.
- ⁴⁸ Menzi L. Behrnd-Klodt, “Balancing Access and Privacy in Manuscript Collections,” *Rights in the Digital Era* (Chicago: Society of American Archivists, 2015), 90.
- ⁴⁹ Behrnd-Klodt, “Balancing Access and Privacy,” 88–89.

- ⁵⁰ Lisl Zach and Marcia Frank Peri, "Practices for College and University Electronic Records Management (ERM) Programs: Then and Now," *The American Archivist* 73, no. 1 (2010): 105–28, <http://www.jstor.org/stable/27802717>.
- ⁵¹ Truman, "Web Archiving Environmental Scan."
- ⁵² Yan Zhang and Barbara M. Wildemuth, "Qualitative Analysis of Content," *Applications of Social Research Methods to Questions in Information and Library Science* (Wesport, Conn.: Libraries Unlimited, 2009), 308–19.
- ⁵³ Thomson, *Preserving Social Media*, 20.
- ⁵⁴ Sellie et al., "Interference Archives," 456; King, "'Say Nothing'"; White, "Minding the Gaps."
- ⁵⁵ Thomson, *Preserving Social Media*, 22–23.
- ⁵⁶ Thomson, *Preserving Social Media*, 24.
- ⁵⁷ NDSA, *Web Archiving Survey Report*, 12; Bailey et al., *Web Archiving*, 18.
- ⁵⁸ Ed Summers, "On Forgetting and Hydration," *Medium*, November 18, 2014, <https://medium.com/onarchivy/on-forgetting-e01a2b95272>.
- ⁵⁹ Truman, "Web Archiving Environmental Scan," 23–24.
- ⁶⁰ Laura Carroll et al., "A Comprehensive Approach to Born-Digital Archives," *Archivaria* 72 (2011): 61–92, <http://pid.emory.edu/ark:/25593/cksgv>.
- ⁶¹ Christine Kim, "Born-Digital and the Virtual Reading Room," *BloggERS!: The Blog of SAA's Electronic Records Section*, February 11, 2016, <https://saaers.wordpress.com/2016/02/11/born-digital-and-in-the-virtual-reading-room/>.
- ⁶² Behrnd-Klodt, "Balancing Access and Privacy," 95.
- ⁶³ Behrnd-Klodt, "Balancing Access and Privacy," 100.
- ⁶⁴ Behrnd-Klodt, "Balancing Access and Privacy," 108.
- ⁶⁵ Documenting the Now, <http://www.docnow.io/>.
- ⁶⁶ Michelle Caswell, Marika Cifor, and Mario H. Ramirez, "'To Suddenly Discover Yourself Existing': Uncovering the Impact of Community Archives," *The American Archivist* 79, no. 1 (2016): 65.

ABOUT THE AUTHOR



Ashlyn Velte is the archivist and assistant professor at the University of Idaho Library Special Collections and Archives. She graduated with an MSLS from the University of North Carolina at Chapel Hill where she also received a certificate in digital curation. While completing her degree, Velte was a Carolina Academic Library associate in the Southern Historical Collection at Wilson Library.