

Where have all the good Tweets gone?

An examination of the information policies at play in Twitter archiving

Georgia Westbrook
Syracuse University, School of Information Studies

Introduction

In 2010, the Library of Congress started archiving every public Tweet after receiving a “starter” archive of Tweets from the first four years of Twitter’s existence (2006-2010); however, the Library could not keep up with the volume of Tweets being produced, so it halted the program at the beginning of 2018.

With changes in data harvesting and mining technologies, the ultimate goal of the project — archiving every Tweet — may be realized in the near future. As a result, it is necessary to examine the policies in place, alongside the issues which informed them, and those policies which may be needed to complete such an undertaking in the future, particularly in regard to privacy, permanence, and user control.

This poster will provide that examination by comparing current information issues and policies from Twitter, the Library of Congress, and other relevant bodies, as well as considering the benefits and challenges in archiving Tweets for both scholars and the public.

Software and Plugins

There are several existing pieces of software and web plugins for archiving Tweets on a smaller scale than the Library of Congress was trying to establish. These options may inform future large scale archiving operations.

Documenting the Now: Twarc is a tool designed to give scholars the opportunity to collect Tweets relevant to their research. “DocNow is an appraisal tool for the web,” allowing archivists to assess what conversations on Twitter might be in need of archiving (DocNow n.d.). The Tweet Catalog and Hydrator work in tandem allowing users both to add their own public Tweet ID sets to a repository and then enabling scholars to “rehydrate” those sets into Tweets with accompanying metadata (Documenting the Now n.d.).

Private tools: Twitter Archiver, created by Digital Inspiration, a private company (Digital Inspiration n.d.) is just one example of the myriad of private plugins that support archiving Tweets; Twitter Archiver allows users to submit a search or hashtag to be tracked; then, Twitter is “polled” hourly and relevant Tweets are put into a Google Sheet (Twitter Archiver n.d.). Since Twitter Archiver is privately owned and not open to edits or changes, it may work for a smaller project, but it does not support the scalability necessary to operate a larger one.



Figure 1. @realDonaldTrump/Screenshot by NPR.

Existing Policies

Twitter policies

- The Library of Congress only archives Tweets that have been publicly tweeted, which aligns with Twitter’s efforts to respect “the privacy and intent of Twitter users” (Lyon & Callahan 2014).
- Several features back up this commitment, including the option to delete a user’s own account and to remove their location data at any time (Lyon & Callahan 2014). Users can also download their own personal archive of Tweets (Twitter n.d.-b).

Library of Congress policies

- The Library set out to acquire all publicly available Tweets, but now acquires them “on a selective basis” with an eye towards public value (Osterberg 2017).
- There are also attendant policies to be considered, including “the creation of access controls to the archive, whether any information should be censored or restricted, and...privacy and user control” (Zimmer 2015).
- Access restrictions were handed down from Twitter when it gifted its archive to the Library; Tweets could be available as soon as six months after they were initially posted, and they would only be accessible to a certain set of staff and researchers (Zimmer 2015).
- Twitter allowed the Library to “dispose” of archived Tweets, but did not articulate the basis on which that decision should be made (Zimmer 2015).

Tweets produced by government bodies and agencies

- The Presidential Records Act of 1978, which was introduced and passed as a result of Richard Nixon’s actions while president, and its updated counterpart, the Presidential and Federal Records Act Amendments of 2014, suggest that presidential Tweets, whether from an official account or from a personal account used while serving as president, are presidential records (Johnson 2018).
- There is not a clear standard governing the archiving of Tweets published by federal agencies. The National Archives and Record Administration published a white paper detailing best practices for Tweets and similar records, but there is not yet a law that makes that, or any consequences, concrete (McCammon 2018).

Professional standards

- The American Library Association Code of Ethics includes several points which are relevant for this project, including a commitment to intellectual freedom, the condemning of censorship, and an interest in promoting privacy and confidentiality; the code also states “we do not advance private interests at the expense of library users, colleagues, or our employing institutions” (American Library Association 2008).
- The Society of American Archivists espouses similar professional pillars in its own Core Values Statement and Code of Ethics, including “social responsibility,” “access and use,” and “privacy” (Society of American Archivists 2008).
- Many information professionals and institutions have written suggested guidelines for agents who may be tasked with enforcing or adhering to these standards while maintaining Twitter archives.

Benefits and Challenges of Archiving Tweets

Benefits

- *Scholarly use:* Documenting the Now is led by scholars from the University of Maryland and the University of Virginia who use social media to chart and preserve social events, including those in Ferguson, Missouri after Michael Brown was shot and killed in 2014 (Documenting the Now n.d.).
- *Self-archiving:* Scholars might also archive their own Tweets, providing a record of their thoughts and interactions on the social media platform, by using tools from Documenting the Now or other plugins available online (Costello & Priem 2011).

Challenges

- *Volume of Tweets:* The change from the Library acquiring *all* Tweets to *selective* Tweets reiterates the fact that, because Tweets are being produced at a huge volume, there are challenges in storing and processing them that need to be examined before it can be done effectively.
- *Subject matter:* Practical challenges include the complexity of subject matter and how it might be articulated in individual Tweets (Zimmer 2015). The Schlesinger Library at the Radcliffe Institute is collecting #MeToo materials; Twitter searches for #MeToo and related terms “have continued to yield around a hundred and fifty thousand Tweets every week, leaving them [researchers] unsure about when to impose a temporal boundary on the archive—or where, in hindsight, historians will locate the end of #MeToo” (Caplan-Bricker 2019).

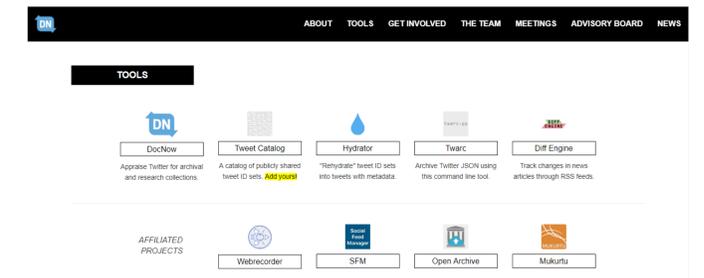


Figure 2. Screenshot from Documenting the Now showing tools and affiliated projects.

Conclusion

After setting out to archive all public Tweets, beginning with a donated archive of Tweets from the first 12 years from Twitter, the Library of Congress quickly realized that the speed at which Tweets were being published was far too fast for effective archiving. Research centers, like Documenting the Now, have taken up the work, albeit with a different plan of attack: rather than collecting Tweets and hosting an archive, Documenting the Now has developed and published open tools that allow scholars to collect their own relevant Tweets. Such self- and scholarly-archiving tools appear to be the immediate future for archiving Tweets, but these tools have the potential to provide for scalability which could inform larger Tweet collection efforts in the future.

Contact

Georgia Westbrook
Syracuse University, School of Information Studies
Syracuse, New York

References

1. American Library Association. (2008). Professional ethics. Retrieved from <http://www.ala.org/tools/ethics>
2. Caplan-Bricker, N. (2019). The challenging of preserving the historical record of #metoo. *The New Yorker*. Retrieved from <https://www.newyorker.com>
3. Costello, K.L. and Priem, J. (2011). Archiving scholars’ tweets. *Society of American Archivists*. Retrieved from <https://www2.archivists.org/sites/all/files/KCFinal.pdf>
4. Digital Inspiration. (n.d.). Retrieved from <https://digitalinspiration.com/>
5. DocNow. (n.d.). Retrieved from <https://github.com/DocNow/docnow>
6. Documenting the Now. (n.d.). Retrieved from <https://www.docnow.io/>
7. George Washington University Libraries. (2018). Social media research ethical and privacy guidelines. Retrieved from <https://gwu-libraries.github.io>
8. Johnson, S. (2018). Donald Trump’s tweets are now presidential records. *Public Radio International*. Retrieved from <https://www.pri.org/>
9. Littman, J. (2019). Twitter’s developer policies for researchers, archivists, and librarians. Retrieved from <https://medium.com/on-archiv>
10. Lyon, G. & Callahan, M. (2014). Honor using intent on Twitter. Retrieved from <http://support.gnip.com/articles/honoring-user-intent-on-twitter.html>
11. McCammon, M. (2018). Can they really delete that? *Slate*. Retrieved from <https://slate.com>
12. Osterberg, G. (2017). Update on the Twitter archive at the Library of Congress. *Library of Congress Blog*. Retrieved from <https://blogs.loc.gov>
13. Raymond, M. (2010). The Library and Twitter: An FAQ. *Library of Congress Blog*. Retrieved from <https://blogs.loc.gov>
14. Society of American Archivists. (2011). SAA Core Values Statement and Code of Ethics. Retrieved from <https://www2.archivists.org/statements>
15. Truman, G. (2016). Web archiving environmental scan. *Harvard Library Report*. Retrieved from <http://nrs.harvard.edu/urn-3:HUL.InstRepos:25658314>
16. Twitter. (n.d.-a). Twitter Developer Platform. Retrieved from <https://developer.twitter.com/en.html>
17. Twitter. (n.d.-b). How to download your Twitter archive. Retrieved from <https://help.twitter.com>
18. Twitter Archiver. (n.d.). Retrieved from <https://chrome.google.com/webstore/category/extensions>
19. Zimmer, M. (2015). The Twitter archive at the Library of Congress: Challenges for information practice and information policy. *First Monday*, 20(7). Retrieved from <http://dx.doi.org/10.5210/firstmonday.2017.5619>