The Problem of Describing Digital Ephemera

by Brinna Michael

In the fall of 2019, preparations began to ramp up for Pitts Theology Library's winter exhibition, *Thy Kingdom Come: American Evangelicalism from George Whitefield to Contemporary Politics*, an exploration of the historical, social, and political complexities of the American Evangelical movement. The goal of curators Brandon Wason and Eric Moore (2019) was to "[problematize] monolithic portrayals of evangelicalism in the media in order to demonstrate the significant role this movement has played within the broader American culture." To accomplish this, they drew on a variety of resources, including original compendiums of George Whitfield's sermons, publications by proponents of Christian fundamentalism, ephemera reflecting the everyday Evangelical life, and books on the state of contemporary American Evangelicalism. In addition to these more traditional, physical representations of embodied history, Wason and Moore also made the decision to include something altogether different for Pitts: a curated selection of tweets to accompany the section on American Evangelicalism and Contemporary Politics.

But why tweets? Pitts Theology Library's Special Collections department specializes in collecting a variety of historical materials largely focused on the Reformation and Wesleyan periods, and it is rare that exhibitions engage so directly with current events. However, Wason and Moore's collaboration presented an opportunity to display a truly contemporary example of the current state of politics and religion, one which could only be captured through social media. Throughout the exhibition's three-month run, Wason, Moore, and Library Director Bo Adams, curated a selection of tweets that were added to a running feed in the gallery, giving a near real-time look into the living Evangelical experience. In fact, the screen on which the feed was running quickly became the section of the exhibition most frequently engaged with by visitors.

In response to the exhibition, some in the Atla community posed the question: amidst cases full of painstakingly preserved and described physical materials, were there plans for preserving and describing the tweets? The short answer is no. A selection of the tweets were included in the official exhibition catalog, but, overall, there was not a plan for any long-term preservation or description of the tweets. This decision was made largely because the displayed tweets constituted an incredibly small and highly curated subset of the social media record of Evangelical engagement with contemporary politics specifically gathered for this exhibition. Rather than attempting to take on a project outside the scope and technical capability of the library, the decision was made not to preserve them outside of their context as a supplemental element of the exhibit itself. Additionally, the preservation of tweets, and social media posts broadly, is an incredibly complex and imprecise process, one that should not be taken on in an ad hoc, one-off fashion.

As the technology we use to communicate and express ourselves continues to develop, it is becoming quite clear that the cultural heritage community must shift our established practices in order to preserve and describe these new forms of cultural communication artifacts. Many are familiar with the often-arduous process of describing physical ephemera or other types of obscure and challenging information artifacts, but digital ephemera present an even more complex undertaking. Tweets and other types of social media posts exemplify the intricacy of such artifacts: interactive by design and ephemeral by nature, it is practically impossible to capture a suitably de-

Brinna Michael is Cataloging and Metadata Librarian at Pitts Theology Library, Emory University.

scriptive context of a collection of tweets through the means available to archivists and librarians now.

WHAT IS A TWEET? WE JUST DON'T KNOW

Twitter currently has a very robust selection of application programming interfaces (APIs), which enable developers to do everything from create and edit tweets and direct messages to retrieve dumps of tweet and public user account metadata in JSON format (Twitter Developer, "Introduction"). These APIs and associated endpoints (applications which perform a specific function, like searching based on a string) were largely designed with business and industry uses in mind, but could potentially be adapted for preservation and description. Still, simple access to the data is not a success in and of itself. There are a number of challenges that pose a particularly tall barrier to any useable process by which to reliably and regularly record and preserve tweets.

The prime example of this is the Library of Congress' attempt to build a Twitter archive through a partnership with Twitter. Beginning in 2010, The Library of Congress and Twitter reached an agreement in which Twitter would provide an archive of public tweets from 2006–10 as well as establish a continual transfer of tweets on an ongoing basis moving forward. That project abruptly stalled, as announced in a white paper by the Library of Congress (2017), which stated that the library "will continue to acquire tweets but will do so on a very selective basis" and that those tweets will be "thematic and event-based, including events such as elections, or themes of ongoing national interest, e.g. public policy." Since then, almost nothing has been heard from the project. Elisabeth Fondren and Meghan Menard McCune (2018), however, provide an excellent analysis of the social, technical, and cultural challenges that impacted the Library of Congress's decision in their article, "Archiving and Preserving Social Media at the Library of Congress: Institutional and Cultural Challenges to Build a Twitter Archive."

Based on Fondren and McCune's observations, let's break down one aspect of the technical issues facing institutions attempting to preserve Twitter content: the challenge of processing and organizing a set of raw data. When a tweet's data is received from Twitter, either by request for a certain account or through an endpoint query, it arrives in the form of metadata encoded in Tweet JSON format. This means that the tweets have been stripped all their original visual and interactive context, calling into question how we even want to define what a tweet "is." Is it the 140 (or 280) characters composing the "text" element of the "Tweet object" (Twitter Developer, "Tweet Object")? Is it the entire set of metadata enclosed in the "Tweet object" itself? Or does it also include the visual rendering of the information as well? This distinction is important when considering the purpose of preserving the information artifact and the method of description to be used for providing access and context. For example, the tweets collected and displayed as part of the *Thy Kingdom Come* exhibition were displayed as they would be seen as part of a Twitter feed, relying on their interlinked visual and textual aspects to engage users in a way that could not be achieved if they were displayed as raw data or even as a more standard metadata "record" view.

TO COLLECT OR NOT TO COLLECT

In addition to such technical issues, there are the ethical and social ones. Privacy is a major concern for social media users, although their behaviors may not align with these concerns (Yerby,

Koohang, and Paliszkiewicz 2019). As cultural heritage institutions, we have the complicated job of trying to balance the responsibility of recording cultural history with respecting the wishes and privacy of creators who may not have intended or wished for their content to be preserved long-term or exposed for research purposes. This is not a new concept. Archives regularly broker access restrictions as part of donor agreements, particularly in regards to living creators and potentially controversial materials. The same considerations should be given when contemplating preserving social media posts:

- What is the purpose of collecting and preserving these posts?

- Are we undermining the creators' privacy and potentially endangering them by collecting and preserving these posts?

- Are we taking appropriate measures to ethically and accurately describe the full context of these posts?

This final question poses perhaps the greatest hurdle to preserving this new wave of digital ephemera. As Fondren and McCune (2018, 41) note, "the permanent and long-term nature of the Library's Twitter Archive has challenged the widespread notion of social media as ephemeral content... [and] may also alter cultural practices on Twitter, as users react to the knowledge that their tweets are now part of history." If the act of collection could affect established cultural practices surrounding social media, how will descriptions and presentation of those collected tweets and posts? Our current methods and systems of description, specifically MARC, have historically been built around concepts of tangible information artifacts, and have thus struggled to be flexible enough to fully contextualize more complex examples of cultural history, such as film, music, performance, physical ephemera, and unpublished works. It only stands to reason that the same level of ingenuity will be called on to fully contextualize something as massive and complex as a social media site.

Additionally, we must remember that the act of curating content being collected by cultural heritage institutions is, in and of itself, a biased act, imposing a particular lens though which cultural and public history will be interpreted for generations to come. The proposal by the Library of Congress (2017) Twitter Archive project to further narrow the collecting parameters moving forward shifts social media into this conversation of collection bias, reinforcing the need for fully fleshed-out methods of contextualization as part of preservation and description.

CONCLUSIONS

The inclusion of tweets in Wason and Moore's exhibition, *Thy Kingdom Come,* exemplified the engaging and dynamic ways in which current and historical methods of cultural communication can interact. While no plans were made to preserve or describe the curated tweets, their inclusion provokes an important discussion for cultural heritage institutions: how are we going to move forward so that future generations will have full, minimally-biased access to the normalized cultural exchanges represented by social media posts and interactions? As with all shifts to the format of cultural exchange, there is no simple answer to this question. While technical challenges present a very concrete hurdle to this goal, we must also carefully consider the ethical and social impact of collecting digital ephemera. Current and widely used methods and standards of description (e.g., MARC, Dublin Core, etc.) cannot accurately define and describe the complex relationships and con-

text of a platform like Twitter. As such, our first goal must be to break free of conventional definitions of information artifacts and clearly answer the question, "What is a tweet?" Until we can define with surety what we are describing, we cannot move forward with designing a descriptive standard.

WORKS CITED

Fondren, Elizabeth and Meghan Menard McCune. 2018. "Archiving and Preserving Social Media at the Library of Congress: Institutional and Cultural Challenges to Build a Twitter Archive." *Preservation, Digital Technology & Culture* 47, no. 2: 33–44. doi.org/10.1516/ptdc-2018-0011.

Library of Congress. 2017. "Update on the Twitter Archive at the Library of Congress."

Twitter Developer. "Introduction to Tweet JSON." *Data Dictionary*.

———. "Tweet Object." *Data Dictionary*.

Wason, Brandon and Eric Moore. 2019. "Introduction." *Thy Kingdom Come: American Evangelicalism from George Whitefield to Contemporary Politics.*

Yerby, Johnathan, Alex Koohang, and Joanna Paliszkiewicz. 2019. "Social Media Privacy Concerns and Risk Beliefs." *Online Journal of Applied Knowledge Management* 7, no. 1: 1–13. doi.org/10.36965/OJAKM.2019.7(1)1-13.