

Elon Musk's buyout of Twitter has placed its user-generated archives in danger

November 23 2022, by Ian Milligan



Credit: Pixabay/CC0 Public Domain

Twitter is in disarray. This is troubling for a platform that comprises no small part of the historical record of today.

While only used by a percentage of Americans ([some 23 percent in 2022](#)) and Canadians ([42 percent of adults in 2018](#)), it has outsized value for sharing information, capturing ongoing events and shaping the cultural conversation.

Twitter's role cannot be underemphasized. In advance of the 2022 American midterm elections, Twitter realized its pivotal role in shaping electoral information meant that its plan to verify anybody who paid

US\$8 could "[sow discord](#)."

Similarly, Twitter is where many turned to information during the opening weeks of the [COVID-19 pandemic](#) and the Ukraine war.

The end of Twitter?

Amidst predictions [of bankruptcy](#) or even wholesale [technical collapse](#), the cultural record of all of these critical moments are [now endangered](#).

This is terrible because the information that our society creates today is tomorrow's historical record. For better or worse, Twitter has been with us throughout the last decade and a half: [election cycles](#), the COVID-19 pandemic (where it has been an exemplar platform for [misinformation](#) and [information](#) alike), and online culture more generally (some tweets have even [become TV shows](#)).

Future historians may be able to learn about these things through media coverage of Twitter, but the ability to access the tweets themselves will be invaluable for historical research. This is doubly true for the spread of information during breaking events, when the platform itself became the main primary source for observers and participants.

Given the centrality of this source, it is hard to believe that it could all disappear. Could it?

Unique vulnerability

Twitter archives take several shapes and sizes. For a time, the most famous one was the Library of Congress's Twitter archive. In 2010, the Library of Congress announced that it would both receive all the text of tweets dating back to 2006 and acquire them going forward.

[Then in December 2017](#), the Library of Congress moved from a collect everything approach to a "selective basis," curating rather [than taking everything](#).

The Internet Archive, a [digital library](#) based in the United States, also collects many Twitter streams, both through its [Wayback Machine](#) and its subscription service [Archive-It](#), where members can choose and curate the accounts that they collect.

Users can go back and look at the suspended (and since reinstated) [@realDonaldTrump account](#), for example. These web archives, however, are targeted: one needs to know the username or particular hashtag that one wants to study.

The Internet Archive's holdings are not at risk, but they are very hard to search and slow to use. To truly unlock the power of Twitter research, more access is required.

At-risk datasets

Fortunately—for now—there is a better way: the [Twitter Application Programming Interface](#) (API). APIs are ways for computer programs to speak to each other. The [Twitter API for Academic Research](#) program allowed researchers to apply for accounts and then design or use programs to create their own collections of both real-time or [historical data](#).

The [DocNow Catalog](#), has a subset of these Twitter collections, and currently has some 142 datasets consisting of over six *billion* tweets, on topics ranging from [#BlackLivesMatter](#) (41 million tweets) to the [2018 American Congressional Election](#) (171 million tweets).

However, to use the API, one needs to agree to the [terms of service](#).

Each [tweet](#) has its own unique number. This means that these datasets do not contain the data, rather they just contain the numbers that are required to *get* the data. In other words, think of it as a library where you could only share the call numbers with other patrons, not the books themselves.

When the API is functional, this makes sense. Every time a search request is made, a dataset is generated. This means that the same search conducted at different moments in time would produce a different dataset. If somebody had deleted their tweet in the meantime, it would not be available for download.

For example, if in 2020 I had tweeted something which was recorded by a researcher but in 2021 decided to delete it, if the dataset was requested in 2022, my tweet would no longer be there.

But if Twitter disappears—or if the API collapses—this data could suddenly become lost. If Twitter was to completely disappear, perhaps scholars could share their original, full datasets. But some of this data may have already been deleted, perhaps due to researchers running out of storage space or facing other institutional or ethical requirements.

We truly are facing the prospect of widespread erasure.

An incalculable loss

The loss of Twitter's 16 years of user-generated content would be a tragedy.

Digital platforms like Twitter are the public town squares of today, unlike more private social media platforms like Facebook. We all have a stake in ensuring its material is preserved: governments, archivists, librarians, historians, activists, among other institutional and private

stakeholders.

Without the Twitter archive, we risk losing important voices from the past. Many of us have experienced elections, protest and the pandemic through [280-character tweets](#). Without these voices, we lose the unique flavor of the tumultuous times we have lived through.

And the next time a platform comes along, it is important for developers to consider how to archive its content for future consideration.

In the meantime, we can [download our own Twitter archives](#). [Several instructional guides have appeared walking users through the process of downloading this data and making it usable](#).

While lacking the preservation power of the Library of Congress, perhaps these digital scrapbooks will one day remind us of the Twitter that was.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Elon Musk's buyout of Twitter has placed its user-generated archives in danger (2022, November 23) retrieved 26 April 2024 from <https://techxplore.com/news/2022-11-elon-musk-buyout-twitter-user-generated.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.