
Social Media Data Preservation in an API-driven World

Amelia Acker

What is this talk about?

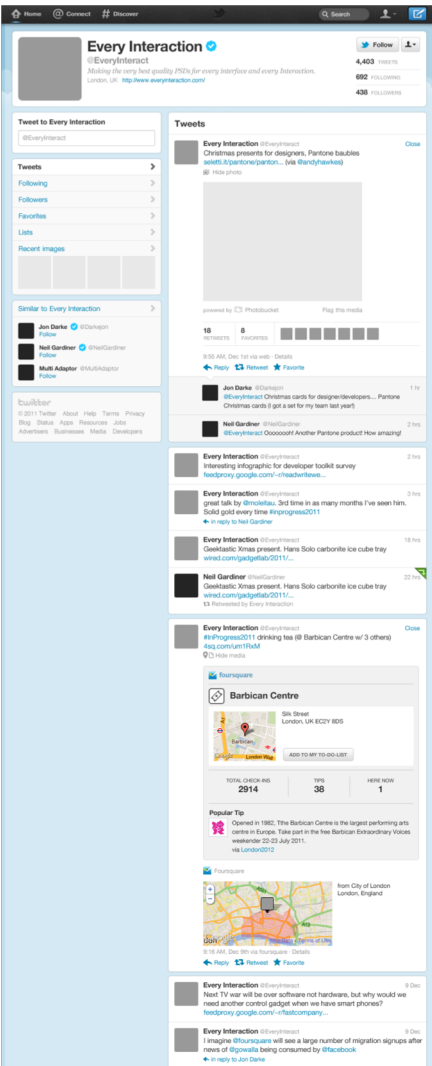
1. Some problems with preserving social media data
 2. What do APIs mean for data persistence?
 3. Some current examples of social media data archives
-

Goals for the audience

APIs are becoming a mark of our digital culture: as access points, they enable data brokers, data infrastructures, and new markets

How does rapid software development and developers' terms of service from social media **platforms** problematize current approaches to digital preservation?

Should we be documenting the APIs themselves?



Social media data: Activity streams

Social media platforms enable users to creation and share a variety of digital objects, many user actions are **locked-in** to platforms

World Wide Web Consortium (W3C) maintains an “Activity Streams” standard defining and describing a JSON-based method for describing a range of user actions from social media

An Activity Stream is a type of metadata associated with digital environments that allows platforms to more consistently manage user data, APIs provide access to these activity streams

Generic Twitter activity stream wireframe.

What's an API?

Application Programming Interface (APIs) are software that specify rules by which software talk to each other

They allow “developers” (users) to articulate queries, limit frequencies and results that are returned

Different APIs allow for purpose driven access and extraction.

Tinder uses Facebook API to match dates

Yelp uses Google Maps API to recommend restaurants

Government Agency uses Twitter API to gather tweets with reports about disasters

Why do APIs matter?

Since the early 2000s, platform APIs have become a source of value and data capital for social media companies, insurance companies, hedge funds, advertising and e-commerce platforms

This API access model assumes data will always be created by users to be collected by platforms to be (re)used as a resource, designating platforms data enablers and service providers

Different users, same API , same terms of service

Even though APIs are one access point, they serve many kinds of users:

- Third party data brokers
- Researchers
- Stewards
- Individual creators

For-profit access models will have different service provision, persistence and digital preservation goals.

—

Existing models

DocNow

Tool suite: Tweet Catalog, Hydrator,
Twarc

Collaborative effort

Web archives, social media tools

Prioritizes and leads in setting
norms for ethical social media data
collection, “*commitment extends to
Twitter's notion of honoring user intent
and the rights of content creators.*”

<https://www.docnow.io/>

Social Feed Manager

Web app that harvests social media data from Twitter, Tumblr, Flickr, and Sina Weibo

University research library

Tool has research, appraisal, and preservation concepts 'baked in'

"Because of the difficulty of getting clarity from Twitter, the ambiguity of Twitter's policies, and our desire to do our work, most of us performing research and archiving with Twitter data make our own private interpretation."

(Littman, 2019)

<https://gwu-libraries.github.io/sfm-ui/>
<https://t.co/ndVr2dWJUY>

Social Media Data Archive

Currently under development at Inter-
university Consortium for Political and
Social Research (ICPSR)

Consortium model

Uses FAIR data principles (all
research objects should be Findable,
Accessible, Interoperable and
Reusable for machines and for
people)

Transformation through metadata
enhancements

<https://www.icpsr.umich.edu/icpsrweb/>
<https://www.nature.com/articles/sdata201618>

Social Science One

Industry-academic partnership

Focuses on the effect of social media
on democracy and elections, with
access to Facebook data

Commission pre-peer review model

Proposed by Gary King and Nathan
Persily

*“mutually incentive-compatible approach
enables academics to analyze and use
the increasingly rich troves of information
amassed by companies to address
societal issues”*

First an API was announced, now
“entirely new systems for securely
storing the data” are being built with
a new team from FB.

<https://socialscience.one/>

Take-aways for the audience

APIs are becoming artifacts of our digital culture

Social media platforms' **terms of service** for developers challenge current approaches to digital preservation

How should we preserve APIs?

thanks!

 @amelia_acker
aacker@ischool.utexas.edu

Acknowledgements



Data&Society

Special thanks to my research collaborator, Dr. Adam Kriesberg @ UMD

This project was made possible in part by:

- Institute of Museum and Library Services Grant RE-07-18-0008-18
- Data & Society Research Institute
- School of Information, UT Austin



References

1. Be a Good Partner to Twitter, Developer Agreement and Policy. Retrieved from <https://developer.twitter.com/en/developer-terms/agreement-and-policy.html>
2. Documenting the Now. (2019). Tweet ID Datasets Catalog. Retrieved from <https://www.docnow.io/catalog/>
3. Hemphill, L., Leonard, S. H., & Hedstrom, M. (2018). Developing a Social Media Archive at ICPSR. In *Proceedings of Web Archiving and Digital Libraries (WADL'18)*. New York, NY. https://doi.org/10.475/123_4
4. King, G., & Persily, N. (2018, August 10). A New Model for Industry-Academic Partnerships. Retrieved from <https://gking.harvard.edu/files/gking/files/partnerships.pdf>
5. Littman, J., Chudnov, D., Kerchner, D., Peterson, C., Tan, Y., Trent, R., ... Wrubel, L. (2016). API-based social media collecting as a form of web archiving. *International Journal on Digital Libraries*, 1–18. <https://doi.org/10.1007/s00799-016-0201-7>
6. Lynch, C. (2017). Stewardship in the “Age of Algorithms.” *First Monday*, 22(12). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/8097>
7. Snell, J. M., & Prodromou, E. (2017). Activity Streams 2.0. Retrieved from <https://www.w3.org/TR/activitystreams-core/>
8. Zubiaga, A. (2018). A longitudinal assessment of the persistence of twitter datasets. *Journal of the Association for Information Science and Technology*, 69(8), 974–984. <https://doi.org/10.1002/asi.24026>