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 archival preservation?
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.
social media as a record

Social media data is an *assemblage* of images, text, metadata, linked content, and engagement data

Shawn Walker @ ASU
“ephemerality of social media data”

Anne Helmond @ University of Amsterdam “platform historiography”

Libby Hemphill @ UMSI + ICPSR
Federated collection building with metadata enhancements

Figure from Shawn Walker (2018) *The Ephemerality of Social Media.*
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.

APIs shape our archival collections, possibilities for access.

This API access model assumes data will always be created by users to be collected by platforms to be (re)used as a resource, for-profit motivations.
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
Interactive recordings of web resources

Captures contexts of user production

Samples of user interface (internal and external) contexts

The need for multiple perspectives on platforms: user experience, developer’s documentation, platform’s commercial decisions

https://webrecorder.io/
**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 extraction-based digital preservation.

How should we preserve APIs?
thanks!

@amelia_acker
aacker@ischool.utexas.edu
Acknowledgements

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