Social Feed Manager & Social Media Archiving

Digital Preservation Coalition
December 6, 2018

Dan Kerchner	@DanKerchner	kerchner@gwu.edu
Laura Wrubel	@liblaura	lwrubel@gwu.edu

Agenda

- Overview of social media data and APIs
- Twitter's API
- Walk-through of Social Feed Manager
- More tools for collecting social media data
- Ethical, legal, and privacy considerations
- Collections at GW Libraries

Social media on the web

Social media as data

id	created_at	user_screen_na	tweet_type	text	hashtags
1067461027867037696	Tue Nov 27 16:51:	theresa_may	retweet	RT @10DowningStreet: "Taken together, the Withdrawal Agreement and the broad terms	
1067426770154967040	Tue Nov 27 14:35:	theresa_may	original	Great to visit the Royal Welsh Winter Fair today to hear from farmers	BackTheBrexitDeal
1067380306368446464	Tue Nov 27 11:31:	theresa_may	original	Baroness Trumpington had a truly extraordinary career – from her work at Bletchley Park	
1067182375619256322	Mon Nov 26 22:24	theresa_may	retweet	RT @10DowningStreet: "Antisemitism and misogyny have no place in this country. Hatre	
1067078616146690049	Mon Nov 26 15:32	theresa_may	retweet	RT @10DowningStreet: WATCH LIVE: PM @Theresa_May updates @H	BackTheBrexitDeal
1067058959935258624	Mon Nov 26 14:14	theresa_may	quote	Many congratulations to @r_j_smith1, selected as the @Conservatives Parliamentary can	
1066760010896736257	Sun Nov 25 18:26:	theresa_may	retweet	RT @10DowningStreet: From the Downing Street steps to the EU Council in Brussels. An i	
1066687749456039938	Sun Nov 25 13:39:	theresa_may	original	I will take this deal back to the House of Commons confident we have	BackTheBrexitDeal
1066672231969947648	Sun Nov 25 12:37:	theresa_may	retweet	RT @10DowningStreet: 40 reasons to back the Brexit deal 👇 [tap to	BackTheBrexitDeal
1066666954893787136	Sun Nov 25 12:16:	theresa_may	retweet	RT @10DowningStreet: WATCH LIVE: PM @Theresa_May makes a sta	BackTheBrexitDeal
1066650176734023680	Sun Nov 25 11:09:	theresa_may	retweet	RT @10DowningStreet: The UK's position on Gibraltar's sovereignty has not changed and	
1066466567700037632	Sat Nov 24 23:00:	theresa_may	original	My letter to the nation. #BackTheBrexitDeal https://t.co/VGzNeeXoc	BackTheBrexitDeal
1066422950008680448	Sat Nov 24 20:06:5	theresa_may	original	I am proud that Gibraltar is British. I will always stand by Gibraltar. https://t.co/NzM1vi8	
1066013570792603649	Fri Nov 23 17:00:1	theresa_may	original	The Brexit deal explained. #BackTheBrexitDeal https://t.co/HvWdvqH	BackTheBrexitDeal
1065980751085281280	Fri Nov 23 14:49:5	theresa_may	original	I've always been clear that getting the right Brexit deal means getting a deal that is right f	
1065917637639708672	Fri Nov 23 10:39:0	theresa_may	retweet	RT @10DowningStreet: Yesterday, PM @Theresa_May visited the James Wigg Practice to	
1065675812878385152	Thu Nov 22 18:38:	theresa_may	retweet	RT @10DowningStreet: PM @Theresa_May spoke in the @HouseofCommons about why	
1065666577729765378	Thu Nov 22 18:01:	theresa_may	retweet	RT @10DowningStreet: PM @Theresa_May met Austrian Chancellor @sebastiankurz at D	
1065621185856237568	Thu Nov 22 15:01:	theresa_may	retweet	RT @10DowningStreet: WATCH LIVE: PM @Theresa_May makes a statement in @Housed	
1065599939902955522	Thu Nov 22 13:36:	theresa_may	original	The British people want this to be settled. They want a good deal that sets us on course f	
1065585241631453184	Thu Nov 22 12:38:	theresa_may	retweet	RT @10DowningStreet: WATCH LIVE: PM @Theresa_May gives a statement on Brexit http	
1065581110896328705	Thu Nov 22 12:21:	theresa_may	retweet	RT @10DowningStreet: Services at @jameswigg help keep people healthy and independe	
1065335026684973057	Wed Nov 21 20:03	theresa_may	original	Our cut to stamp duty is just one of the steps we've taken to fix the housing market. I'm	

```
{
    "entities": {
        "hashtags": [],
        "symbols": [],
        "urls": [
            {
                "display_url": "facebook.com/TheresaMayForP…",
                "expanded_url":
"https://www.facebook.com/TheresaMayForPM/posts/1349767708373332",
                "indices": [
                    118,
                    141
                ],
                "url": "https://t.co/TqTTz0hlIz"
            }
        ],
        "user_mentions": []
    },
    "favorite_count": 825,
    "favorited": false,
    "full_text": "Today I launch my campaign to be our next PM. Please read &amp;
share my vision of a country that works for everyone: https://t.co/TqTTz0hlIz TM",
    "geo": null,
    "id": 748427251713187840,
```

APIs
and
JSON

Collecting social media content (typically) involves

using the platform's **API**

to get posts as **JSON**

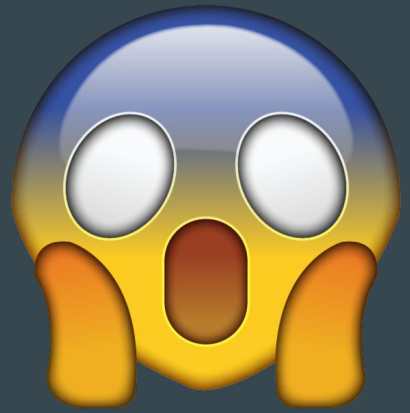
What's an API?

- Short for Application Programming Interface
- Intended for ***software*** to interact with a website
  - vs. a web interface, designed for ***people*** to interact with a web site.
- API calls consist of:
  - **request**: `http://an.api.com/somerequest?foo=15`
  - **response**: structured data, e.g., XML or JSON

Why not just use Web Scraping?

Why not just use Web Scraping?

- It's hard
- It's fragile
- It's incomplete



Or maybe traditional
Web Archiving?

But using the API...

Using an API for collecting social media data

- Generally gives you exactly what the platform stores.
- Can give you useful slices of data you can't get by any amount of scraping.
- Gives you social media data in structured format, which makes it easy to analyze as data.

JSON - JavaScript Object Notation

- `{ key: value, key: value… }`
- keys are strings
- a value may be:
  - string - in quotes: `"GW"`
  - number
  - boolean - `true` or `false`
  - another JSON object
  - array (denoted by square brackets `[ ]`) of JSON objects
  - `null`

JSON example

```
{'title': 'Social Media Archiving',
 'logistics':
    {'datetime_utc': '2018-12-06 14:30:00',
     'duration_hrs': 1,
     'location': 'Webex'},
 'awesome': true,
 'wine_list': [],
 'presenters': ['Laura Wrubel',
                'Dan Kerchner']
}
```

Tweets are JSON too

- Twitter's guide to the structure of a tweet: https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object
- Let's look at an example...

GW Arts & Sciences
@gwucolumbian

Following

Congratulations to CCAS alums Sally Nuamah, BA '11, and Tara Dorfman, BA '11, who were both named to Forbes 2019 30 Under 30 lists! #GWCCAS #GWU

11:29 AM - 2 Dec 2018

3 Retweets **9** Likes

3 9

https://twitter.com/gwucolumbian/status/1069267199318216704

```
{
  created_at: "Sun Dec 02 16:29:00 +0000 2018",
  id: 1069267199318216700,
  id_str: "1069267199318216704",
  full_text: "Congratulations to CCAS alums Sally Nuamah, BA '11, and Tara Dorfman
  named to Forbes 2019 30 Under 30 lists! #GWCCAS #GWU https://t.co/kPWyKrJ9Ux",
  truncated: false,
+ display_text_range: […],
- entities: {
  - hashtags: [
    - {
        text: "GWCCAS",
      - indices: [
          132,
          139
        ]
      },
    - {
        text: "GWU",
      - indices: [
          140,
          144
        ]
      }
    ],
    symbols: [ ],
    user_mentions: [ ],
    urls: [ ],
  - media: [
    - {
        id: 1069267196357091300,
        id_str: "1069267196357091328",
      - indices: [
          145,
          168
        ],
        media_url: "http://pbs.twimg.com/media/DtbM2ZDXoAAMvQ7.jpg",
        media_url_https: "https://pbs.twimg.com/media/DtbM2ZDXoAAMvQ7.jpg",
        url: "https://t.co/kPWyKrJ9Ux",
        display_url: "pic.twitter.com/kPWyKrJ9Ux",
        expanded_url: "https://twitter.com/gwucolumbian/status/10692671993182
        type: "photo",
```



```
  in_reply_to_status_id: null,
  in_reply_to_status_id_str: null,
  in_reply_to_user_id: null,
  in_reply_to_user_id_str: null,
  in_reply_to_screen_name: null,
- user: {
    id: 54639174,
    id_str: "54639174",
    name: "GW Arts & Sciences",
    screen_name: "gwucolumbian",
    location: "Washington, D.C.",
    description: "Official Twitter account for the Columbian College of Arts an
    Sciences at The George Washington University. Home of the Engaged Liberal
    Arts.",
    url: "https://t.co/g0ys0T59mJ",
  - entities: {
    - url: {
      - urls: [
        - {
            url: "https://t.co/g0ys0T59mJ",
            expanded_url: "http://columbian.gwu.edu",
            display_url: "columbian.gwu.edu",
          - indices: [
              0,
              23
            ]
          }
        ]
      },
    - description: {
        urls: [ ]
      }
    },
    protected: false,
    followers_count: 4322,
    friends_count: 608,
    listed_count: 136,
    created_at: "Tue Jul 07 18:49:10 +0000 2009",
    favourites_count: 4202,
    utc_offset: null,
    time_zone: null,
    geo_enabled: true,
    verified: false,
    statuses_count: 9273,
    lang: "en",
```

Most useful API methods for collecting tweets

- User timeline: GET statuses/user_timeline
- Search: GET search/tweets
- Filter stream: POST statuses/filter
- Sample stream

More: https://developer.twitter.com/en/docs

Geotagging

- When posting a tweet:
  - Geotagging is opt-in. Only ~2% geotagged.
  - Lat, long or place name (e.g., DC or Middle Earth)
- API support:
  - Search API: Limit to a specified distance of a lat, long.
  - Filter Stream: Limit to a bounding box.

More:
https://gwu-libraries.github.io/sfm-ui/posts/2017-04-12-geographic-collecting

Social Feed Manager walkthrough

Social Feed Manager

Social Feed Manager empowers researchers and archivists to build collections of social media data from multiple platforms.

Log In

If you have not created an account yet, then please sign up first.

Username

Username

Password

Password

Forgot your password?

Remember me

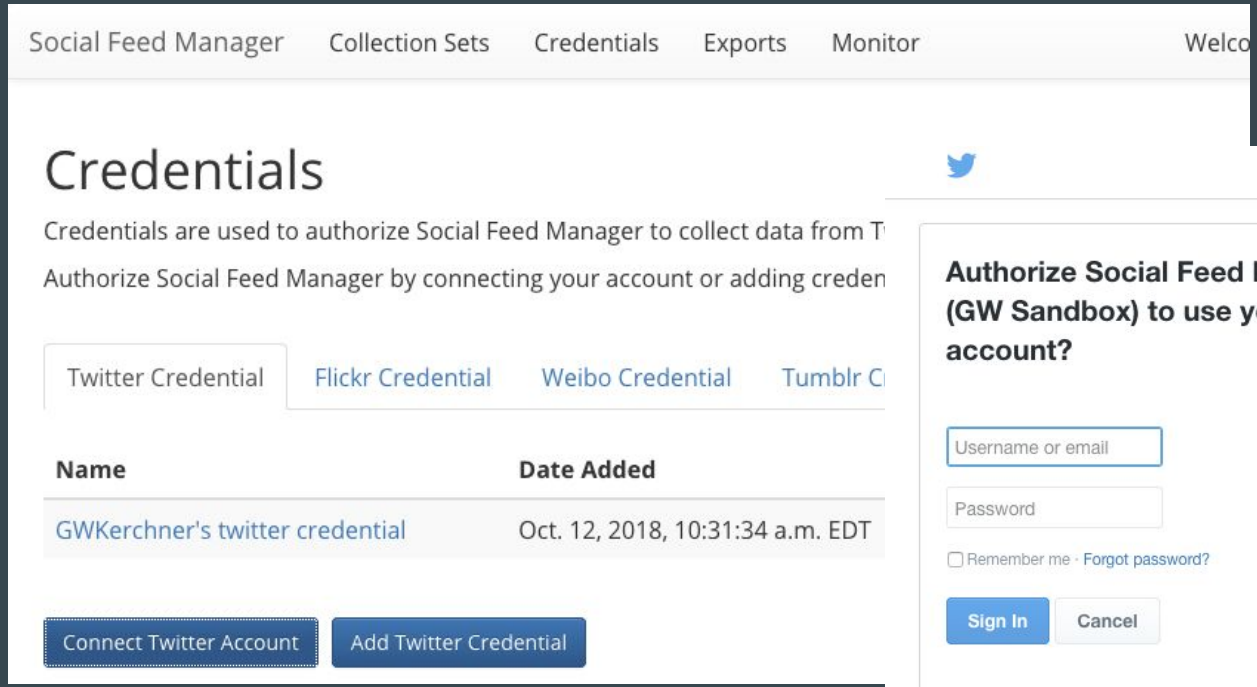
Log in

Social Feed Manager empowers researchers and archivists to create collections of social media data from Twitter, Tumblr, Flickr, and Sina Weibo. Read more about Social Feed Manager here.

Collecting and using data from social media platforms is subject to those platforms' terms (Twitter, Flickr, Sina Weibo, Tumblr), as you agreed to them when you created your social media account. Social Feed Manager respects those platforms' terms as an application (Twitter, Flickr, Sina Weibo, Tumblr).

Social Feed Manager provides data to you for your research and academic use. Social media platforms' terms of service **generally do not allow republishing of full datasets,** and you should refer to their terms to understand what you may share. Authors typically retain rights and ownership to their content.

In addition to respecting the platforms' terms, as a user of Social Feed Manager and data collected within it, it is your responsibility to consider the ethical aspects of collecting and using social media data. Your discipline or professional organization may offer guidance. Here are some ethical and privacy guidelines you may want to consider.


Social Feed Manager
Collection Sets
Credentials
Exports
Monitor
Welco
Credentials
Credentials are used to authorize Social Feed Manager to collect data from T
Authorize Social Feed Manager by connecting your account or adding creden
Twitter Credential
Flickr Credential
Weibo Credential
Tumblr C
Name
Date Added
GWKerchner's twitter credential
Oct. 12, 2018, 10:31:34 a.m. EDT
Connect Twitter Account
Add Twitter Credential

Sign up for Twitter ›
Authorize Social Feed Manager (GW Sandbox) to use your account?
Username or email
Password
Remember me · Forgot password?
Sign In
Cancel
This application will be able to:
• Read Tweets from your timeline.
• See who you follow.
Will not be able to:
• Follow new people.
• Update your profile.
• Post Tweets for you.
• Access your direct messages.
• See your email address.
• See your Twitter password.
Social Feed Manager (GW Sandbox)
By GW Libraries
go.gwu.edu/sfm
Social Feed Manager, GW Sandbox

Social Feed Manager | Collection Sets | Credentials | Exports | Monitor | Welcome, kerchner2

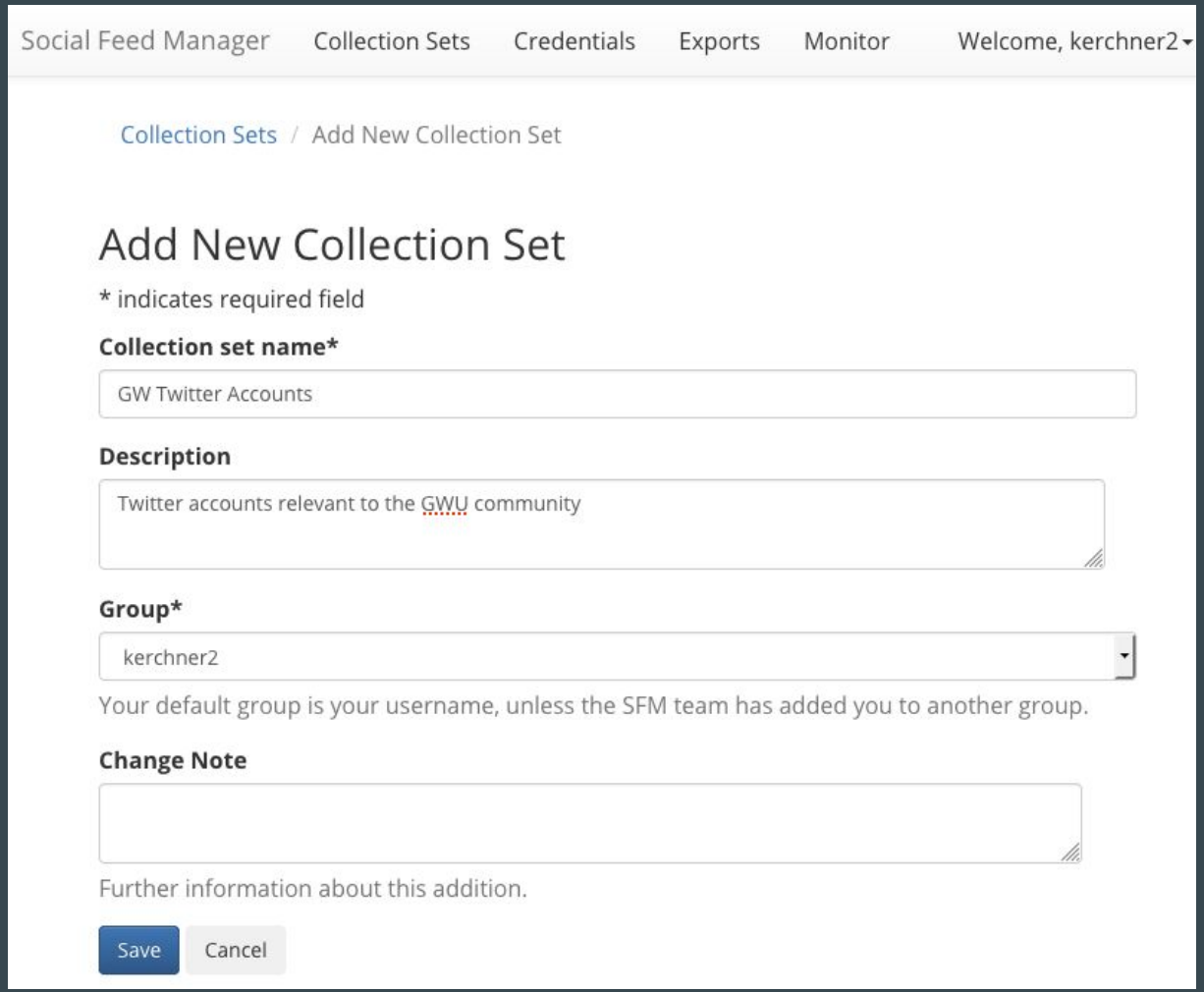
Collection Sets

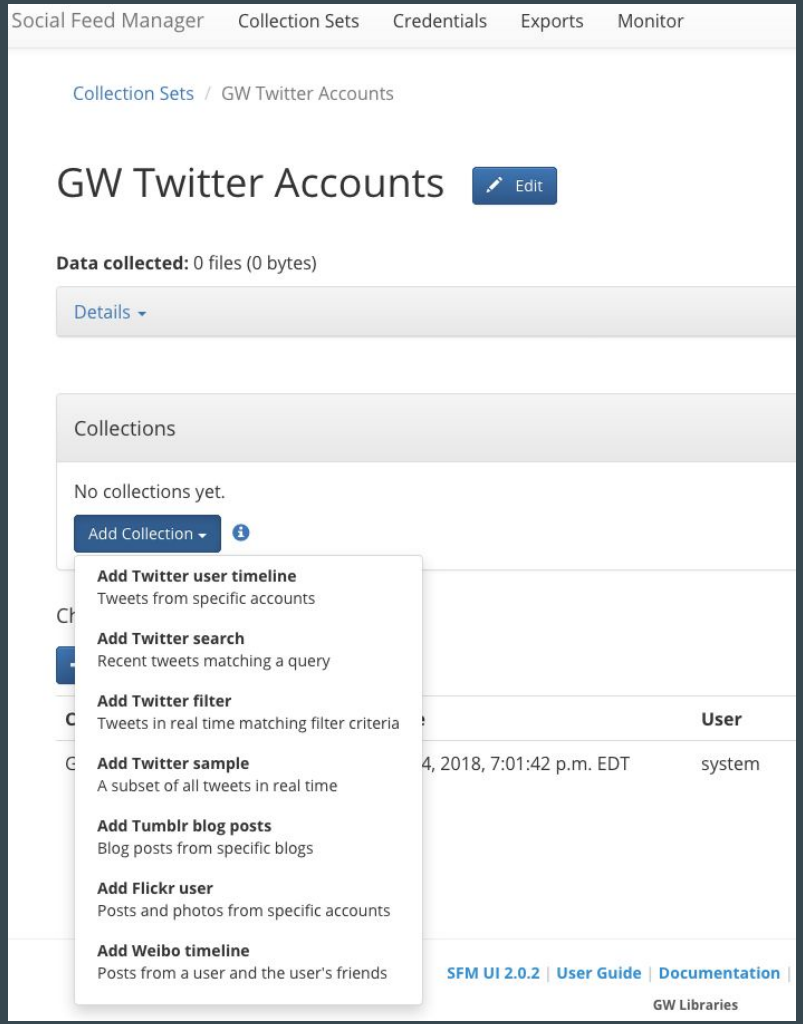
A collection set is a group of collections around a particular topic or theme. Collections sets are active when there is at least one active collection within them. Collection sets are inactive when all collections have been deactivated and are no longer harvesting.

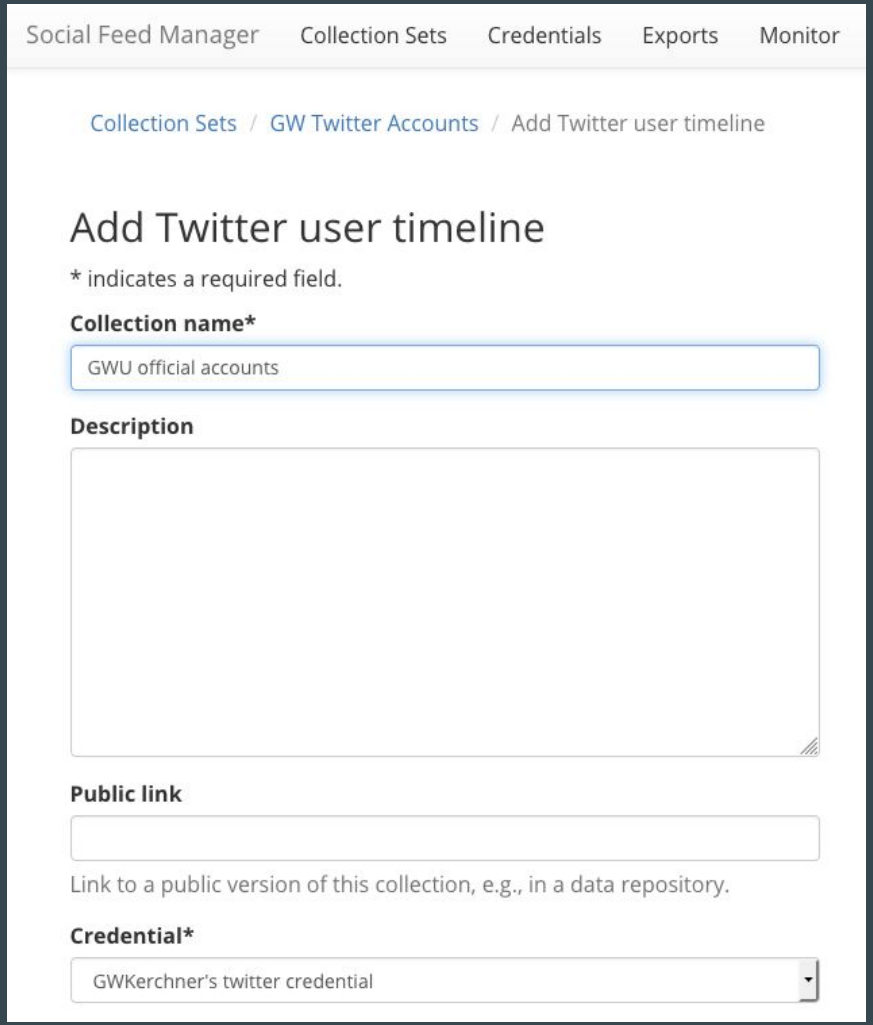
Active 0 | Shared 4

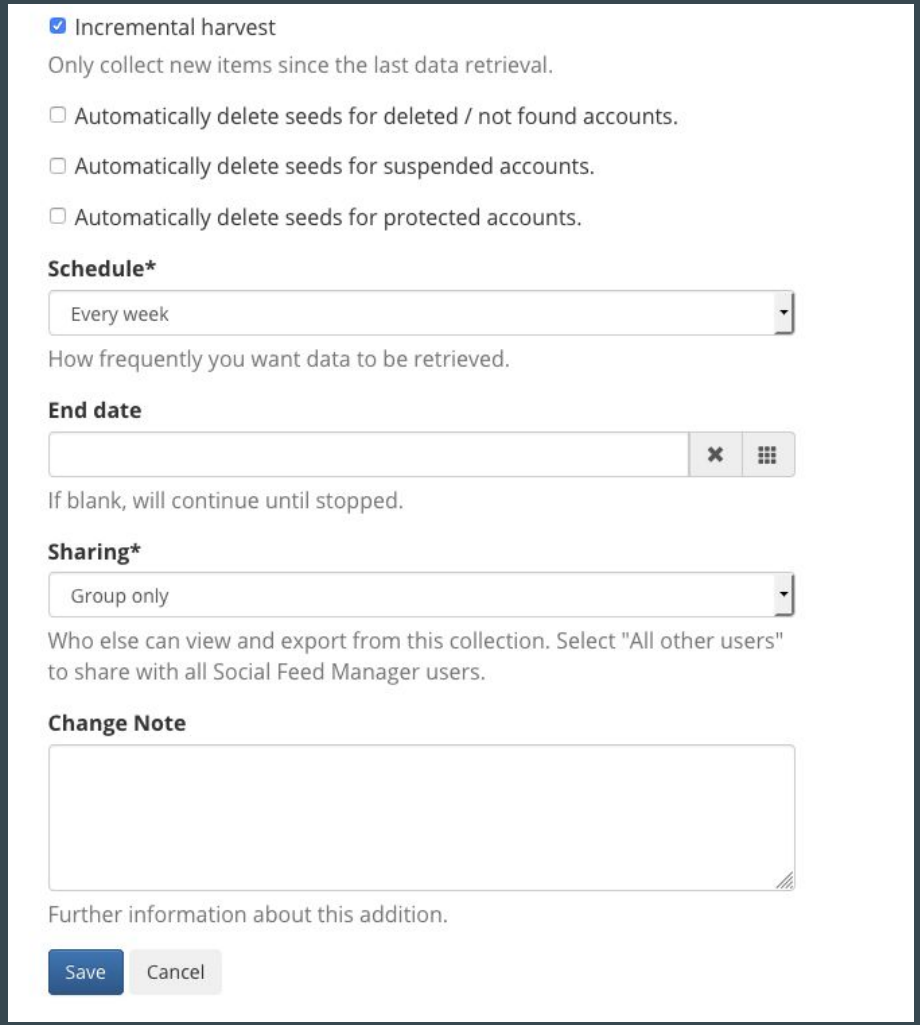
No corresponding collection sets.

Add Collection Set


Social Feed Manager
Collection Sets
Credentials
Exports
Monitor
Welcome, kerchner2
Collection Sets / Add New Collection Set
Add New Collection Set
* indicates required field
Collection set name*
GW Twitter Accounts
Description
Twitter accounts relevant to the GWU community
Group*
kerchner2
Your default group is your username, unless the SFM team has added you to another group.
Change Note
Further information about this addition.
Save
Cancel


Social Feed Manager
Collection Sets
Credentials
Exports
Monitor
Collection Sets / GW Twitter Accounts
GW Twitter Accounts
Edit
Data collected: 0 files (0 bytes)
Details
Collections
No collections yet.
Add Collection
Add Twitter user timeline
Tweets from specific accounts
Add Twitter search
Recent tweets matching a query
Add Twitter filter
Tweets in real time matching filter criteria
Add Twitter sample
A subset of all tweets in real time
Add Tumblr blog posts
Blog posts from specific blogs
Add Flickr user
Posts and photos from specific accounts
Add Weibo timeline
Posts from a user and the user's friends
User
4, 2018, 7:01:42 p.m. EDT
system
SFM UI 2.0.2 | User Guide | Documentation |
GW Libraries


Social Feed Manager
Collection Sets
Credentials
Exports
Monitor
Collection Sets / GW Twitter Accounts / Add Twitter user timeline
Add Twitter user timeline
* indicates a required field.
Collection name*
GWU official accounts
Description
Public link
Link to a public version of this collection, e.g., in a data repository.
Credential*
GWKerchner's twitter credential


Incremental harvest
Only collect new items since the last data retrieval.
Automatically delete seeds for deleted / not found accounts.
Automatically delete seeds for suspended accounts.
Automatically delete seeds for protected accounts.
Schedule*
Every week
How frequently you want data to be retrieved.
End date
If blank, will continue until stopped.
Sharing*
Group only
Who else can view and export from this collection. Select "All other users" to share with all Social Feed Manager users.
Change Note
Further information about this addition.
Save
Cancel

Social Feed Manager | Collection Sets | Credentials | Exports | Monitor | Welcome, kerchner2

Collection Sets / GW Twitter Accounts / GWU official accounts

GWU official accounts

Twitter user timeline

Turn on | Deactivate | Edit | Export

New collection added. You can now add seeds.

At least 1 active seed must be added before harvesting can be turned on.

Data collected: 0 files (0 bytes)

Details

Seeds

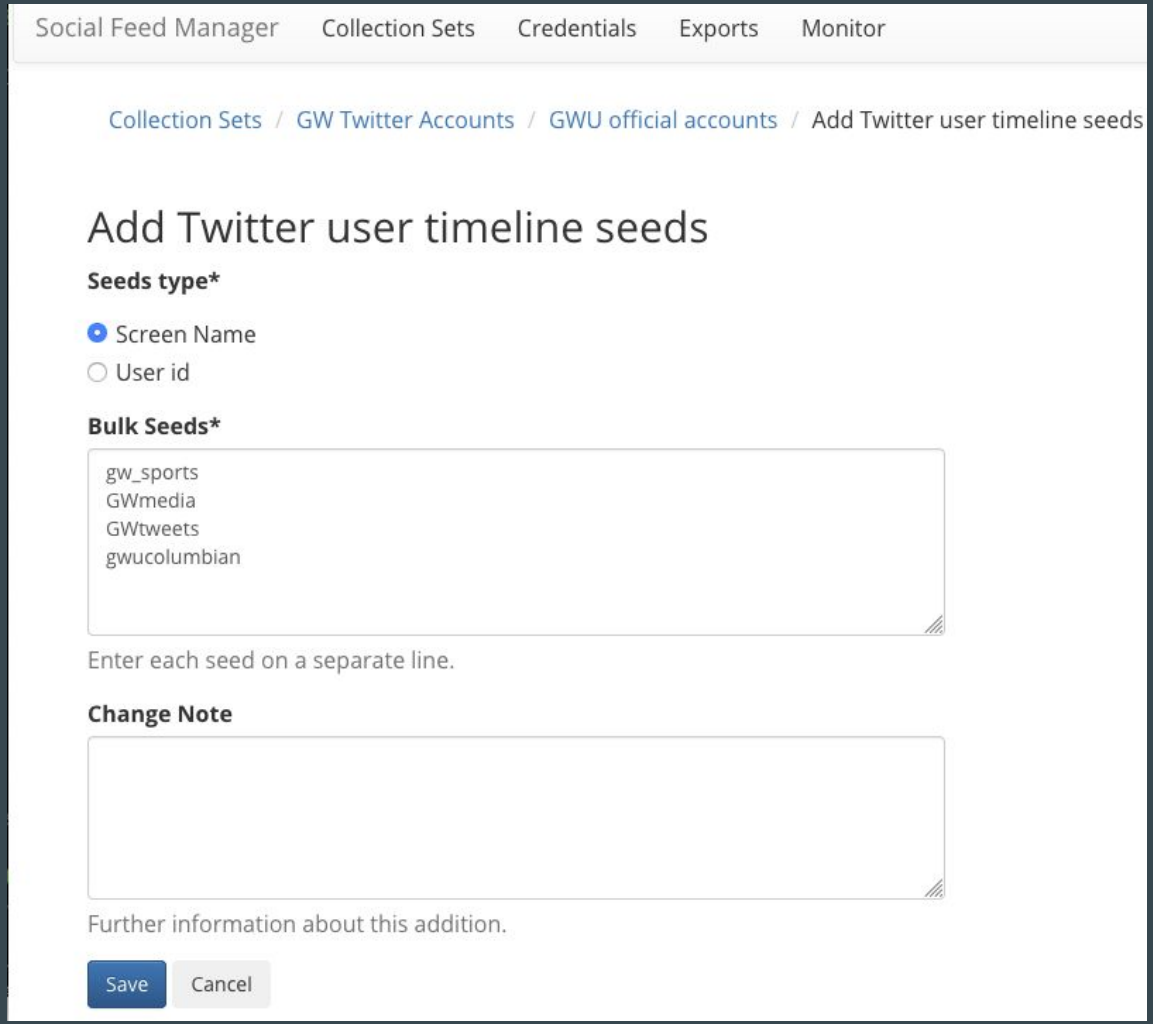
Download seed list

Active | Deleted

Search

Link	Twitter accounts	User ID	Messages

Add Seed | Bulk Add Seeds


Social Feed Manager
Collection Sets
Credentials
Exports
Monitor
Collection Sets / GW Twitter Accounts / GWU official accounts / Add Twitter user timeline seeds
Add Twitter user timeline seeds
Seeds type*
Screen Name
User id
Bulk Seeds*
gw_sports
GWmedia
GWtweets
gwucolumbian
Enter each seed on a separate line.
Change Note
Further information about this addition.
Save
Cancel

Collection Sets / GW Twitter Accounts / GWU official accounts

GWU official accounts

Turn off | Deactivate | Edit | Export

Twitter user timeline

Collection is active. Turn off to edit.

Next harvest at Dec. 4, 2018, 5:29:18 p.m. EST

Data collected: 1 file (4.7 MB)

Stats:

- tweets: 12,865

Details

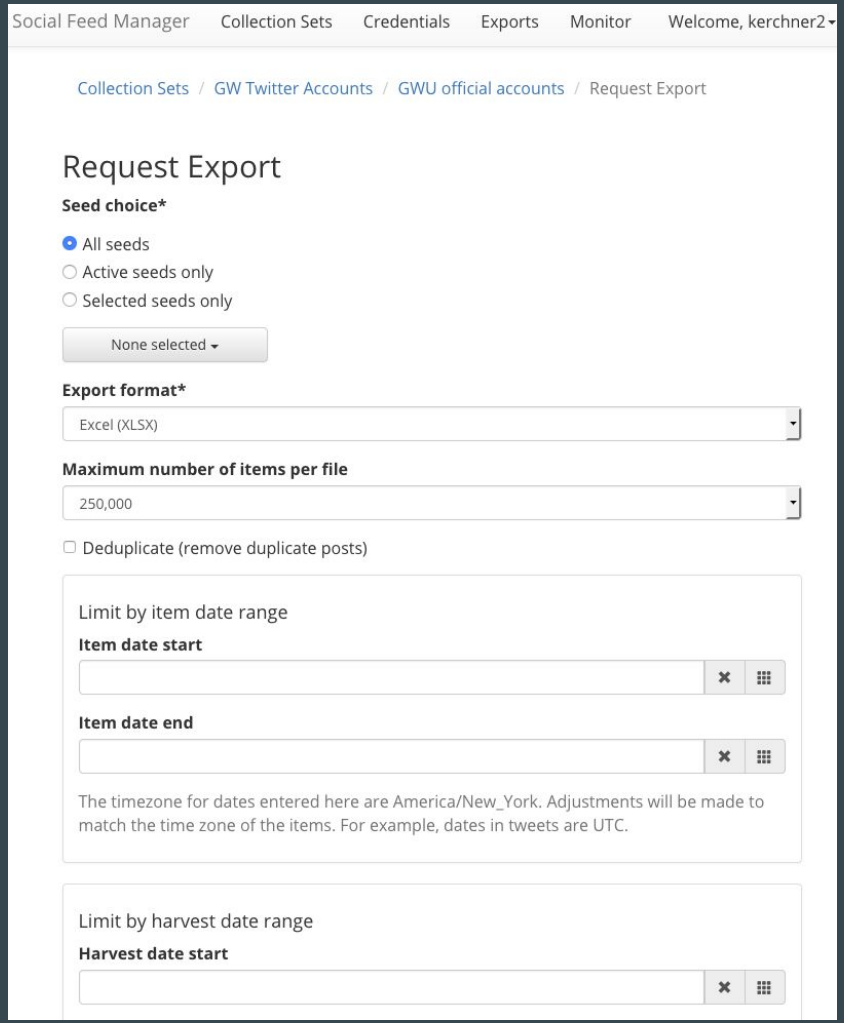
Seeds

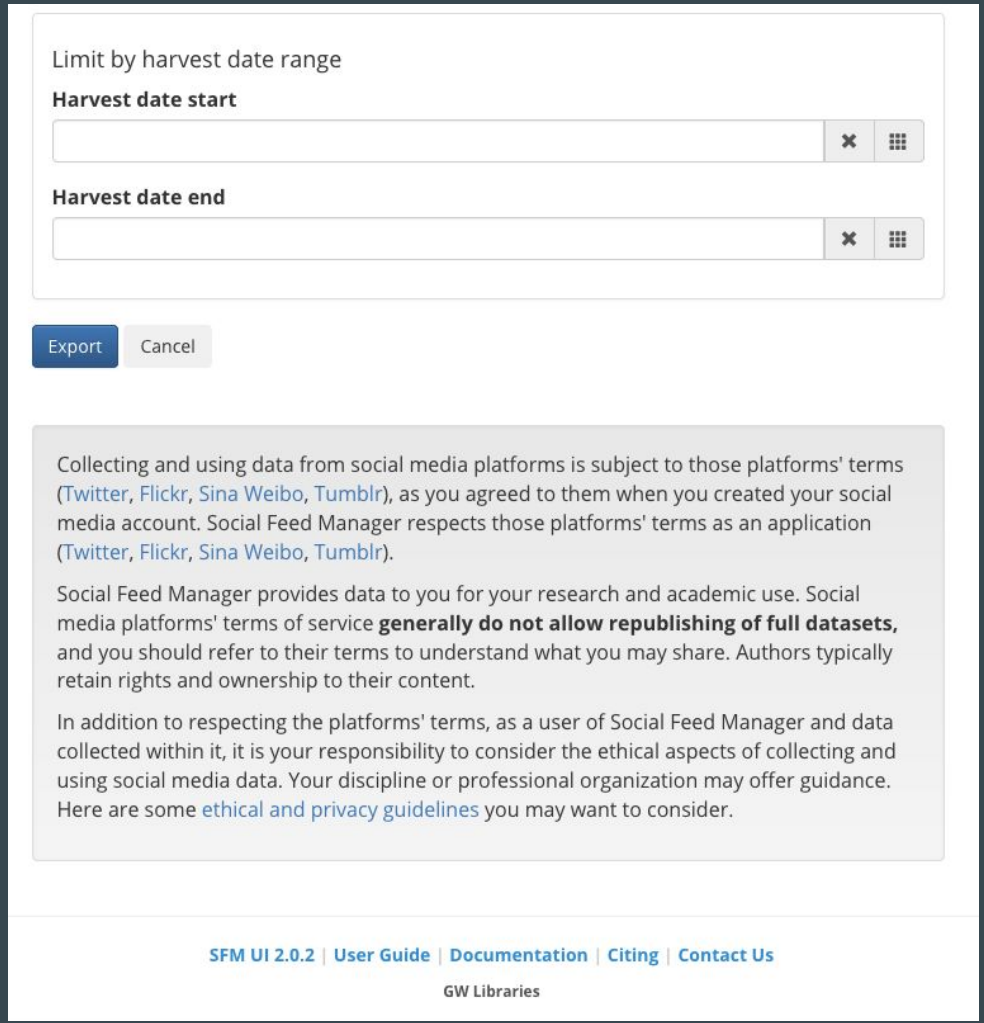
Download seed list

Active 4 | Deleted

Search

Link	Twitter accounts	User ID	Messages
	GWmedia	59863790	
	gw_sports	54627173	
	GWtweets	28101965	
	gwucolumbian	54639174	


Social Feed Manager
Collection Sets
Credentials
Exports
Monitor
Welcome, kerchner2
Collection Sets / GW Twitter Accounts / GWU official accounts / Request Export
Request Export
Seed choice*
All seeds
Active seeds only
Selected seeds only
None selected
Export format*
Excel (XLSX)
Maximum number of items per file
250,000
Deduplicate (remove duplicate posts)
Limit by item date range
Item date start
Item date end
The timezone for dates entered here are America/New_York. Adjustments will be made to match the time zone of the items. For example, dates in tweets are UTC.
Limit by harvest date range
Harvest date start


Limit by harvest date range
Harvest date start
Harvest date end
Export
Cancel
Collecting and using data from social media platforms is subject to those platforms' terms (Twitter, Flickr, Sina Weibo, Tumblr), as you agreed to them when you created your social media account. Social Feed Manager respects those platforms' terms as an application (Twitter, Flickr, Sina Weibo, Tumblr).
Social Feed Manager provides data to you for your research and academic use. Social media platforms' terms of service generally do not allow republishing of full datasets, and you should refer to their terms to understand what you may share. Authors typically retain rights and ownership to their content.
In addition to respecting the platforms' terms, as a user of Social Feed Manager and data collected within it, it is your responsibility to consider the ethical aspects of collecting and using social media data. Your discipline or professional organization may offer guidance. Here are some ethical and privacy guidelines you may want to consider.
SFM UI 2.0.2 | User Guide | Documentation | Citing | Contact Us
GW Libraries

Export files for GWU official accounts

Filename	Size
6b0da4c9bb0e4087a48c8793effe0bab-README.txt	2.0 KB
6b0da4c9bb0e4087a48c8793effe0bab_001.xlsx	2.8 MB

See the data dictionary for more information about the fields in the export.

See the guidance on citing SFM and datasets.

Status: Success

Format: xlsx

Selected seeds: All seeds

Details

Collecting and using data from social media platforms is subject to those platforms' terms (Twitter, Flickr, Sina Weibo, Tumblr), as you agreed to them when you created your social media account. Social Feed Manager respects those platforms' terms as an application (Twitter, Flickr, Sina Weibo, Tumblr).

Social Feed Manager provides data to you for your research and academic use. Social media platforms' terms of service **generally do not allow republishing of full datasets,** and you should refer to their terms to understand what you may share. Authors typically retain rights and ownership to their content.

In addition to respecting the platforms' terms, as a user of Social Feed Manager and data collected within it, it is your responsibility to consider the ethical aspects of collecting and using social media data. Your discipline or professional organization may offer guidance. Here are some ethical and privacy guidelines you may want to consider.

id	tweet_url	created_at	parsed_cre	user_scree	text	tweet_type	coordin	hashtags	media	urls	favorite_cc
106970559017435	https://twitte	Mon Dec 0	2018-12-03	gwucolumb	Kyrah Altman, a senior majoring in human services and social justice, was a	original		CCASOutFr	https://twitter.com/gv		0
106967441805249	https://twitte	Mon Dec 0	2018-12-03	gwucolumb	RT @CorcoranGW: Professor Maria del Carmen Montoya, featured in a rec	retweet					0
106934420665575	https://twitte	Sun Dec 02	2018-12-02	gwucolumb	The new #GWCCAS International Buddy Program wants to help internation	original		GWCCAS G	https://twi	https://bit.	1
106926719931821	https://twitte	Sun Dec 02	2018-12-02	gwucolumb	Congratulations to CCAS alums Sally Nuamah, BA '11, and Tara Dorfman, BA	original		GWCCAS G	https://twitter.com/gv		9
106896923730842	https://twitte	Sat Dec 01	2018-12-01	gwucolumb	A #GWCCAS junior and her peers got the opportunity to see Michelle Obam	original		GWCCAS O	https://twi	https://bit.	3
106888619715248	https://twitte	Sat Dec 01	2018-12-01	gwucolumb	Last week, @CorcoranGW hosted Washington D.C.'s annual National Portfo	original		GWU GWC	https://twi	https://bit.	3

in_reply_to	in_reply_to	in_reply_to	lang	place	possibly_se	retweet_co	retweet_or_que	retweet_or_quo	retweet_or	source	user_id	user_creat	user_defau	user_descr	user_favou	user_follov	user_frien
			en		FALSE	0				<a href="ht	54639174	Tue Jul 07 1	FALSE	Official Twi	4202	4322	608
			en			1	1069666074830	CorcoranGW	216058845	<a href="ht	54639174	Tue Jul 07 1	FALSE	Official Twi	4202	4322	608
			en		FALSE	0				<a href="ht	54639174	Tue Jul 07 1	FALSE	Official Twi	4202	4322	608
			en		FALSE	3				<a href="ht	54639174	Tue Jul 07 1	FALSE	Official Twi	4202	4322	608
			en		FALSE	0				<a href="ht	54639174	Tue Jul 07 1	FALSE	Official Twi	4202	4322	608
			en		FALSE	1				<a href="ht	54639174	Tue Jul 07 1	FALSE	Official Twi	4202	4322	608

Social Feed Manager | Collection Sets | Credentials | Exports | Monitor | Welcome, kerchner

Collection Sets

A collection set is a group of collections around a particular topic or theme. Collections sets are active when there is at least one active collection within them. Collection sets are inactive when all collections have been deactivated and are no longer harvesting.

Active 22 | Inactive 13 | Shared 23 | Other Active 109 | Other Inactive 12

Name	Collections	Date Added	Groups
115th U.S. Congress	3 collections	Jan. 27, 2017, 10:47:12 a.m. EST	GW Libraries Scholarly Technology Group
2017-2020 Federal Term	2 collections	Jan. 20, 2017, 11:26:54 a.m. EST	GW Libraries Scholarly Technology Group
Alaska Earthquake	2 collections	Nov. 30, 2018, 1:37:25 p.m. EST	GW Libraries Scholarly Technology Group
China Anti-Corruption	48 collections	June 29, 2016, 11:49:14 p.m. EDT	CEAL Grant
Climate change	1 collection	Sept. 21, 2017, 8:11:54 a.m. EDT	GW Libraries Scholarly Technology Group
Corcoran	1 collection	March 29, 2017, 5:17:54 p.m. EDT	GW Libraries Scholarly Technology Group
Foreign leaders	1 collection	March 11, 2018, 9:43:54 p.m. EDT	GW Libraries Scholarly Technology Group
Governors	1 collection	March 11, 2018, 9:34:03 p.m. EDT	GW Libraries Scholarly Technology Group

Infrastructure

- SFM:
  - Production
  - Sandbox: For instruction and experimentation.
  - Amazon Web Services instances for special uses.
- Access is restricted to GW network.
- Currently using 5TB out of 6TB storage

Infrastructure: Challenges

- Skills
- Storage Space
- Updates / Keeping up with API changes

More tools for collecting social media data

- DocNow tools:
  - twarc: command-line tools for collecting from Twitter, with a suite of utilities for visualization and processing.
  - Hydrator: desktop app which will get the full tweet metadata for a list of tweet identifiers
    (watch for more in the second phase of their Mellon grant)
- TAGS: Google Sheet for quick and dirty collecting.
- Commercial tools

Great list from IIPC of more tools.

DocNow Slack, for getting advice from the community.

Commercial social media services

- Limitations of Twitter service providers may include:
  - A focus on marketing and business intelligence, not academic research.
  - Restricted data exports, often 50K tweets per day.
  - Analytics that are “black box” and may not be fit for academic research and archiving.
- Purchasing tweets can be pricey. Price depends on number of tweets and length of the time period.

Techniques for collection development

- Search Twitter website.
- Identify hashtags with DocNow App.
- Extract accounts from Twitter lists.
- Extract hashtags / accounts from collections.
- Rely on existing lists
  - USNPL
  - U.S. Digital Registry (U.S. federal agencies)
  - Wikipedia
  - Ballotpedia

Ethical considerations

Social media data comes from people

- Consider impact of your work on the creator of the social media.
- Do not have creator's permission for research.
- Impact on creator is balanced against public good of your research.
- Requires judgement call.

Social media research ethical and privacy guidelines at GW

"'Participant' Perceptions of Twitter Research Ethics." Casey Fiesler, Nicholas Proferes, *Social Media + Society.* First published March 10, 2018. doi.org/10.1177/2056305118763366

Data collecting

Be thoughtful collecting social media of:

- Vulnerable individuals (e.g., minors, social activists)
- Sensitive or harmful topics (e.g., questionable behavior, mental illness)
- Geography-based collecting

Publishing

- When possible, get permission from creator for quotes from social media.
- Do not rely on anonymizing posts.
- Respect a user’s deleting content.

Data sharing

- Get familiar with platform terms of use.
  - Don’t republish full datasets
  - Share in accordance with terms (e.g., tweet ids only)
  - Consider copyright
- Sharing summary statistics is usually OK.

Documenting the Now White Paper on Ethics

1. Archivists should engage and work with the communities they wish to document.
2. Documentation efforts must go beyond what can be collected without permission from the web and social media.
3. Archivists should follow social media platforms' terms of service where they are congruent with the values of the communities they are attempting to document.
4. When possible, archivists should apply traditional archival practices such as appraisal, collection development, and donor relations to social media and web materials.

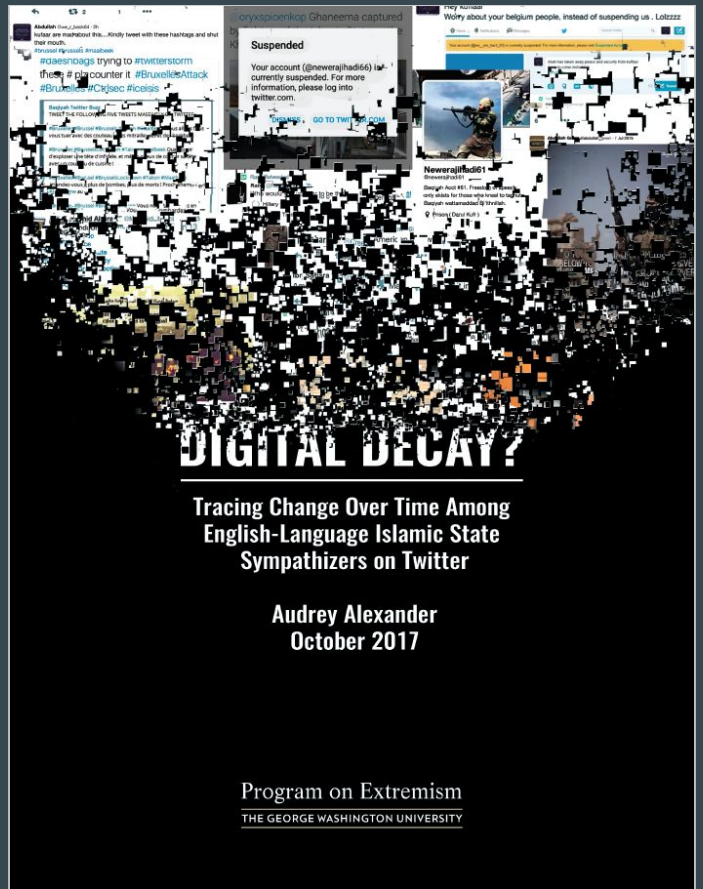
Collecting at GW

- Performed primarily by SFM team.
- Attempt to collect on topics related to GW's research strengths.
- Have experimented collaborating with subject librarians.
- Consultations rather than self-service.
- Informed by "Building Social Media Archives: Collection Development Guidelines"

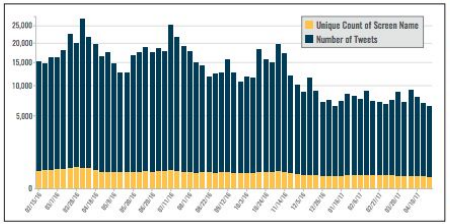
Example collections

- GWU-related collecting (collecting from our own community, official accounts)
- 2016 and 2018 U.S. election:s general hashtags, Democrat hashtags, GOP hashtags, candidate accounts
- Hurricanes and natural disasters
- U.S. government accounts
- News organizations
- Climate change
- Members of Congress

Social media research



Tweet Frequency and Unique Screen Names By Week

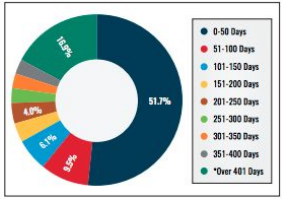


(Figure 5) This graph shows how the relationship between unique screen names and tweet frequency per week changed over the course of the 63-week period. As discussed in the method chapter, this graph, like several others in the study, uses square root in the y-axis to more clearly represent the relationship between the two variables.

of activity' is quantified by the number of days between an account's first and last tweet. Twitter's API does not discern the date or time at which the company suspends accounts, nor does it identify accounts that were created and then subsequently abandoned by their respective users. Consequently, this measurement allows the study to grasp the chronological span of sympathizers that actively use the platform to share content. While overwhelmingly skewed by outliers, the average lifespan for this sample of English-language pro-IS accounts on Twitter was 251 days. It is critical to note, however, that dispersion of lifespan is highly concentrated (see Figure 6). Approximately 51.7 percent of accounts did not remain active longer than 50 days. On the other hand, however, a substantive portion of accounts lasted over a year, suggesting that Twitter's attempts to detect and suspend pro-IS account may be missing some long-term users. One possible explanation for long-standing users relates to the data collection method, as researchers are more likely to identify accounts the longer they are open.[15] Ultimately, accounts that opted to leave the platform are likely included in this breakdown, although multiple factors- including the threat of suspension- likely affect user activity in this regard.

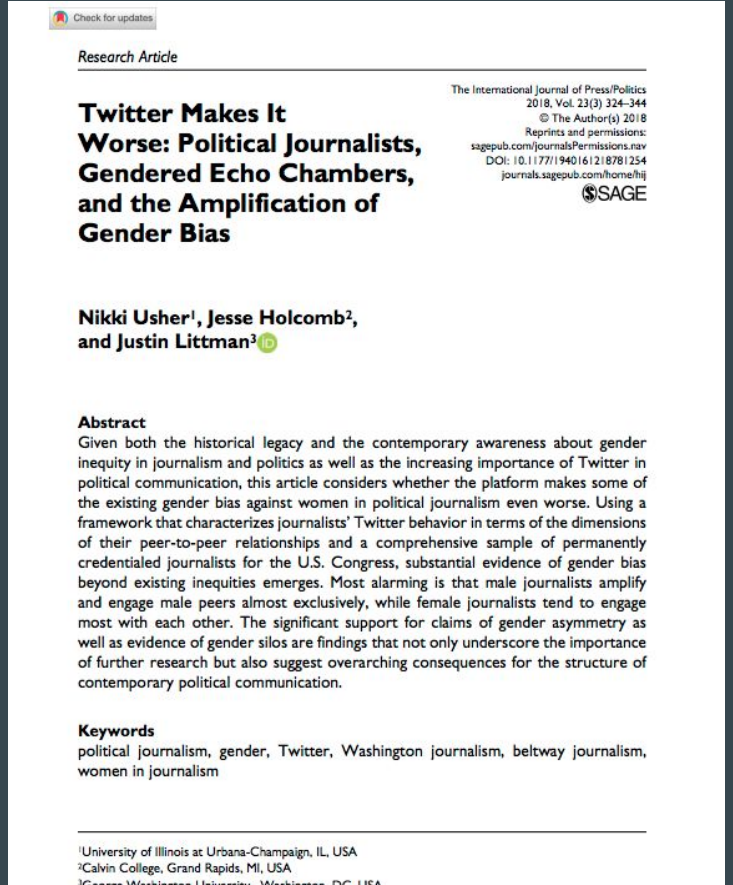
In order to maintain their presence on Twitter, some English language IS sympathizers appeared to have created multiple accounts at the same time to avoid shutdowns. On February 17, 2016, for example, four separate accounts were fashioned from a core handle,[16] possibly from the same individual. One account (@erhabi35) survived only eight days, whereas another account (@Erhabi39) stayed active for 62 days. Although the study attempted to annotate cases where the same individual controlled multiple accounts, as the trend is common, quantitative figures are generally not reliable due to the relative anonymity Twitter affords users. It is hard to ascertain whether users that demonstrate similar behavioral patterns are simply individuals attempting to inoculate their digital presence against suspensions or are

Duration of Account Activity



(Figure 6) This chart depicts the duration of account activity, meaning the number of days a pro-IS Twitter account was active, and displays the breakdown in percentages.

Social media research



Check for updates

Research Article

The International Journal of Press/Politics
2018, Vol. 23(3) 324–344
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1940161218781254
journals.sagepub.com/home/hij
SAGE

Twitter Makes It Worse: Political Journalists, Gendered Echo Chambers, and the Amplification of Gender Bias

Nikki Usher[1], Jesse Holcomb[2], and Justin Littman[3]

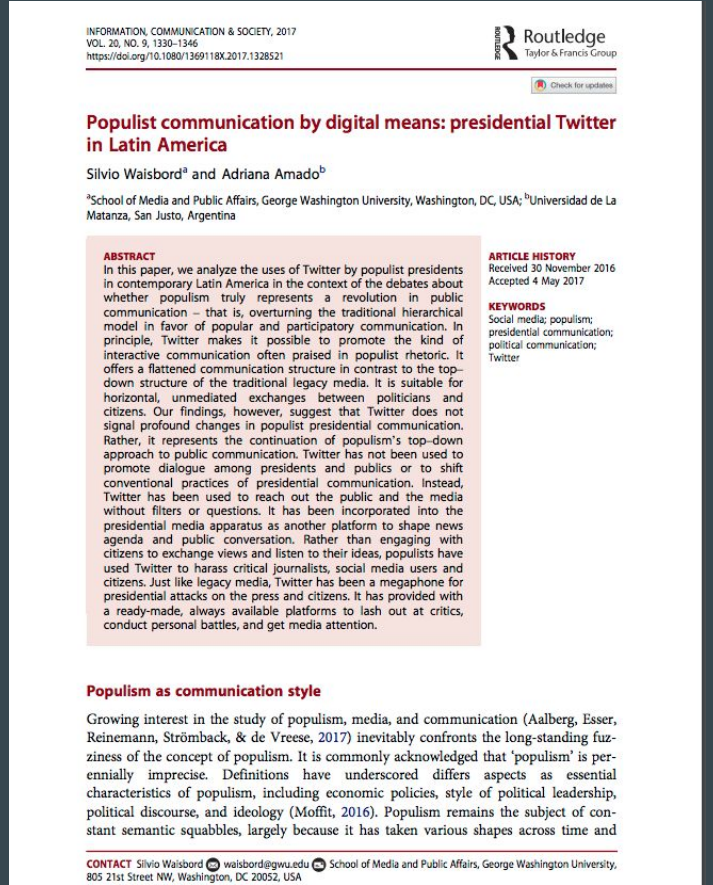
Abstract

Given both the historical legacy and the contemporary awareness about gender inequity in journalism and politics as well as the increasing importance of Twitter in political communication, this article considers whether the platform makes some of the existing gender bias against women in political journalism even worse. Using a framework that characterizes journalists' Twitter behavior in terms of the dimensions of their peer-to-peer relationships and a comprehensive sample of permanently credentialed journalists for the U.S. Congress, substantial evidence of gender bias beyond existing inequities emerges. Most alarming is that male journalists amplify and engage male peers almost exclusively, while female journalists tend to engage most with each other. The significant support for claims of gender asymmetry as well as evidence of gender silos are findings that not only underscore the importance of further research but also suggest overarching consequences for the structure of contemporary political communication.

Keywords

political journalism, gender, Twitter, Washington journalism, beltway journalism, women in journalism

[1]University of Illinois at Urbana-Champaign, IL, USA
[2]Calvin College, Grand Rapids, MI, USA
[3]George Washington University, Washington, DC, USA



INFORMATION, COMMUNICATION & SOCIETY, 2017
VOL. 20, NO. 9, 1330–1346
https://doi.org/10.1080/1369118X.2017.1328521

Routledge
Taylor & Francis Group

Check for updates

Populist communication by digital means: presidential Twitter in Latin America

Silvio Waisbord[a] and Adriana Amado[b]

[a]School of Media and Public Affairs, George Washington University, Washington, DC, USA; [b]Universidad de La Matanza, San Justo, Argentina

ABSTRACT

In this paper, we analyze the uses of Twitter by populist presidents in contemporary Latin America in the context of the debates about whether populism truly represents a revolution in public communication – that is, overturning the traditional hierarchical model in favor of popular and participatory communication. In principle, Twitter makes it possible to promote the kind of interactive communication often praised in populist rhetoric. It offers a flattened communication structure in contrast to the top–down structure of the traditional legacy media. It is suitable for horizontal, unmediated exchanges between politicians and citizens. Our findings, however, suggest that Twitter does not signal profound changes in populist presidential communication. Rather, it represents the continuation of populism's top–down approach to public communication. Twitter has not been used to promote dialogue among presidents and publics or to shift conventional practices of presidential communication. Instead, Twitter has been used to reach out the public and the media without filters or questions. It has been incorporated into the presidential media apparatus as another platform to shape news agenda and public conversation. Rather than engaging with citizens to exchange views and listen to their ideas, populists have used Twitter to harass critical journalists, social media users and citizens. Just like legacy media, Twitter has been a megaphone for presidential attacks on the press and citizens. It has provided with a ready-made, always available platforms to lash out at critics, conduct personal battles, and get media attention.

ARTICLE HISTORY
Received 30 November 2016
Accepted 4 May 2017

KEYWORDS
Social media; populism; presidential communication; political communication; Twitter

Populism as communication style

Growing interest in the study of populism, media, and communication (Aalberg, Esser, Reinemann, Strömback, & de Vreese, 2017) inevitably confronts the long-standing fuzziness of the concept of populism. It is commonly acknowledged that 'populism' is perennially imprecise. Definitions have underscored differs aspects as essential characteristics of populism, including economic policies, style of political leadership, political discourse, and ideology (Moffit, 2016). Populism remains the subject of constant semantic squabbles, largely because it has taken various shapes across time and

CONTACT Silvio Waisbord waisbord@gwu.edu School of Media and Public Affairs, George Washington University, 805 21st Street NW, Washington, DC 20052, USA

Questions?

Dan Kerchner @DanKerchner kerchner@gwu.edu
Laura Wrubel @liblaura lwrubel@gwu.edu

Social Feed Manager
sfm@gwu.edu
go.gwu.edu/sfm