

Using Web Archives to Document Social Movements and Disinformation - Practice, Ethics, and Challenges

Dr. Michele C. Weigle

Professor, Department of Computer Science
Web Science and Digital Libraries (WS-DL) Group
Old Dominion University



<https://bit.ly/weigle-lucy-2023>

Featuring work done in collaboration with

Web Science and Digital Libraries (WS-DL) Research Group at ODU



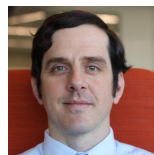
PhD Students and Alumni



Dr. Michael L. Nelson



Sawood Alam
(PhD 2020)



Mat Kelly
(PhD 2019)



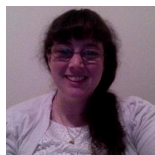
Hany SalahEldeen
(PhD 2015)

MS Alumni



Nauman Siddique
(MS 2020)

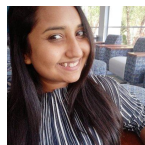
REU Students (undergrad)



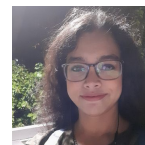
Lesley Frew



Kritika Garg



Himarsha Jayanetti



Tarannum Zaki



Haley Bragg

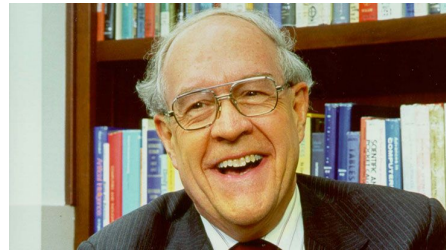


Caleb Bradford



Computer scientists are toolsmiths

The Computer Scientist as **Toolsmith**



Frederick P. Brooks, Jr.

Frederick P. Brooks, Jr. 1996. The computer scientist as toolsmith II. Commun. ACM 39, 3 (March 1996), 61-68,
<http://www.cs.unc.edu/~brooks/Toolsmith-CACM.pdf>



UNC
DEPARTMENT OF
COMPUTER SCIENCE

The Web holds our stories

ESPN.com

March 13, 2020

TOP EVENTS ▼

NCAAM

Canceled
SJU
CREI


Canceled
CLEM
FSU

Canceled
MINN
IOWA

Canceled
BUT
PROV

Canceled
OKST
KU

ESPN



NCAA tournaments canceled over coronavirus

The NCAA has called off its men's and women's basketball tournaments as part of a complete cancellation of all remaining spring and winter championships.

5h • ESPN News Services


What we know and don't know after NCAA tournament cancellations

CNN.com

March 14, 2020

CNN

Trump declares national emergency



LIVE UPDATES

The declaration comes as there is only one US state left without coronavirus cases

Get info: Send your stories | Podcast | Newsletter | Alerts

Trump says he takes no responsibility for early testing failures

Analysis: Trump's appalling, blame-shifting briefing

See exec's response to Trump's attempt to shake hands

But webpages can change or disappear


ESPN.com

April 14, 2023

TOP EVENTS ▾ NBA 7:00 PM ET TNT 9:30 PM ET ESPN RBC Heritage ESPN+ Round 2 - In Progress 1. M. Hubbard -8 MLB >

CHI 40-42 OKC 40-42 PGA Tour MIN 42-40 MIA 44-38


ESPN



Debating biggest need vs. best value in Round 1: We made two picks at all 31 NFL draft slots

Fill a need? Best player available? We played out both strategies at all 31 first-round slots in the 2023 draft.

ESPN 4h • Matt Miller and Jordan Reid



Stacking the draft's top cornerbacks: How they shut down wideouts


CNN.com

April 14, 2023

Audio Live TV Log In

LIVE UPDATES: Ukraine Documents leak Stock market TRENDING: Fort Lauderdale Dalai Lama NRA

Pentagon document leak suspect charged



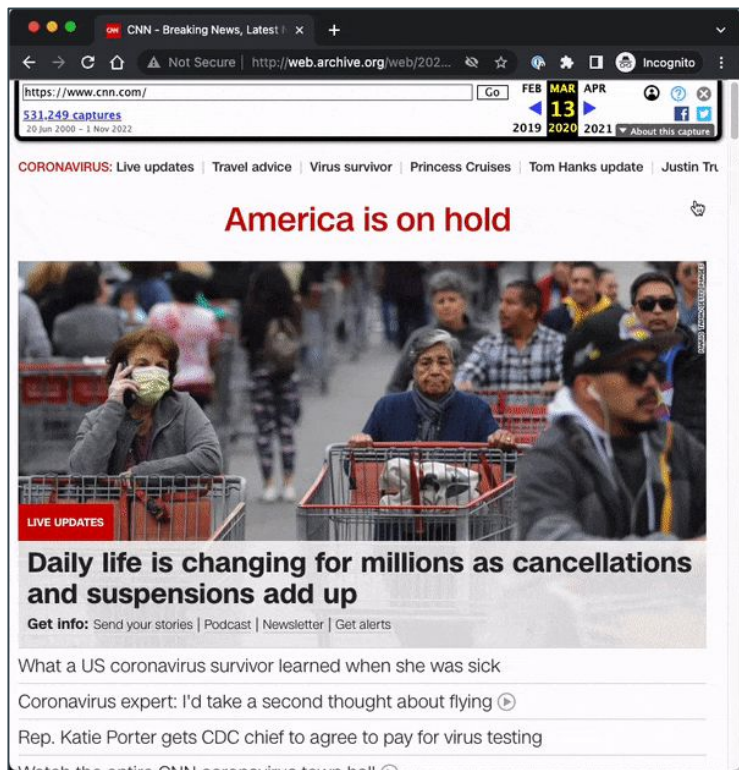
LIVE UPDATES

The 21-year-old Air National Guardsman appeared in federal court in Boston a day after his arrest at his Massachusetts home

Teixeira entered court in shackles; his hands were uncuffed

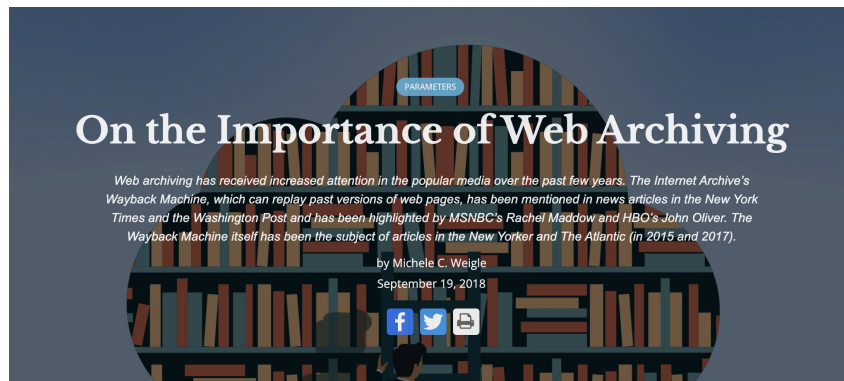
Man arrested in connection with intel leak was a loner fascinated

Maybe it's archived?



[CNN.com archived on March 13, 2020](https://web.archive.org/web/20200313112222/https://www.cnn.com/)

More than just saving a screenshot, web archives strive to capture and allow replay of the *entire contents of a web page*, including its source HTML and embedded images, stylesheets, and JavaScript source code.



Michele C. Weigle, "On the Importance of Web Archiving", *SSRC Parameters*, Sep 2018, <https://items.ssrc.org/parameters/on-the-importance-of-web-archiving/>

Web archives are essential for studying recent history and culture



Ian Milligan (University of Waterloo), <https://twitter.com/NetPreserve/status/1141321443373920256>

There are several public web archives



and others...

https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives


How do webpages get archived?

Anatomy of a Web Page

Browser: About Me - Michele C. Weigle | x | +
Address Bar: https://weiglemc.github.io
Incognito

Michele C. Weigle @ ODU | Publications | Teaching | Students | Talks | Schedule

About Me

 **Dr. Michele C. Weigle**
Professor of Computer Science

Research Interests: web science, social media, web archiving, digital preservation, information visualization (see some student infovis projects in my [infovis gallery](#))

Location: Norfolk, VA
Affiliation: Old Dominion University
Website:

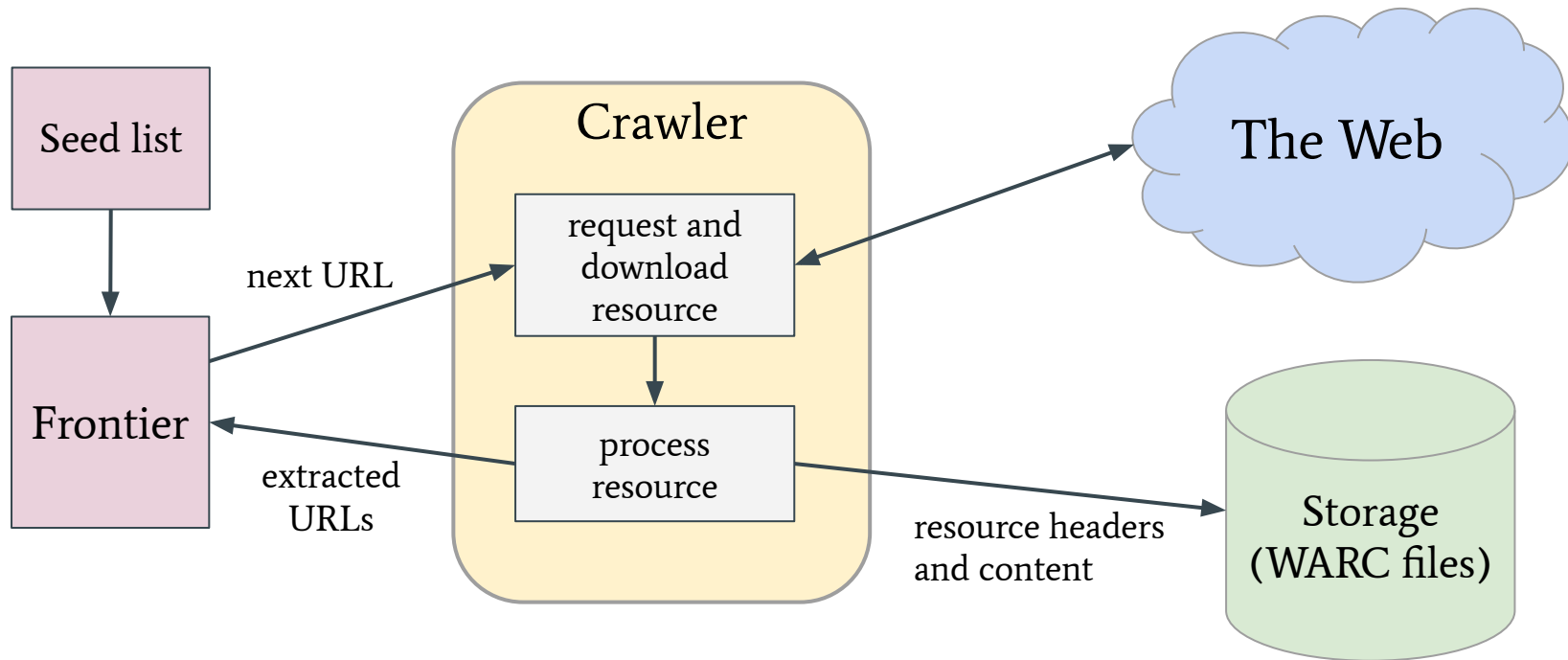
Links: [On the importance of web archiving](#), [SSRC Parameters](#), [ODU Faculty Women in STEM page](#), [bio.](#)

Annotations: CSS, JavaScript

<https://www.cs.odu.edu/~mweigle/>

How do webpages get archived?

Web Crawling Process



On-Demand Web Archiving



Add Seeds
Enter one seed URL per line below to add them to this collection.
☒ https://craftycocktails.com/pages/mixology-101/

Access:
Frequency:
Seed Type:

www.archive-it.org

<https://web.archive.org>

**Save Page Now**



<https://archive.today>

My url is alive and I want to archive its content



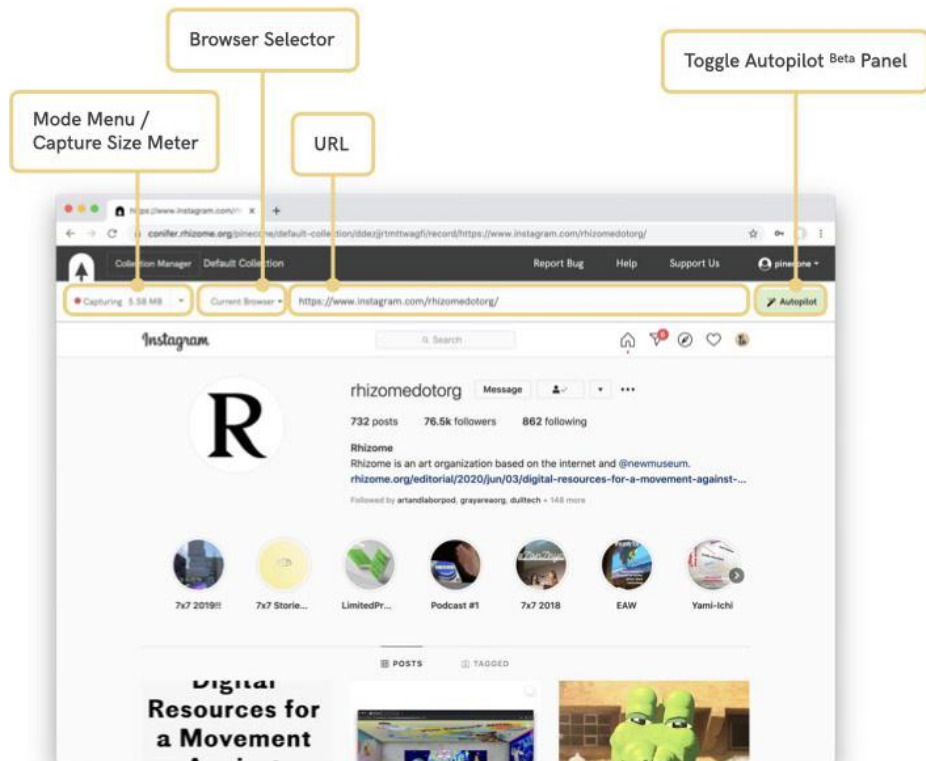
New Capture

Add to collection:
Select browser:
Session settings:

browser	version	release	OS	capabilities
Chrome	v76	2019-08-05	linux	flash, autopilot
Firefox	v68	2019-07-09	linux	flash, autopilot

<https://conifer.rhizome.org>

Webrecorder and Conifer



<https://guide.conifer.rhizome.org/>

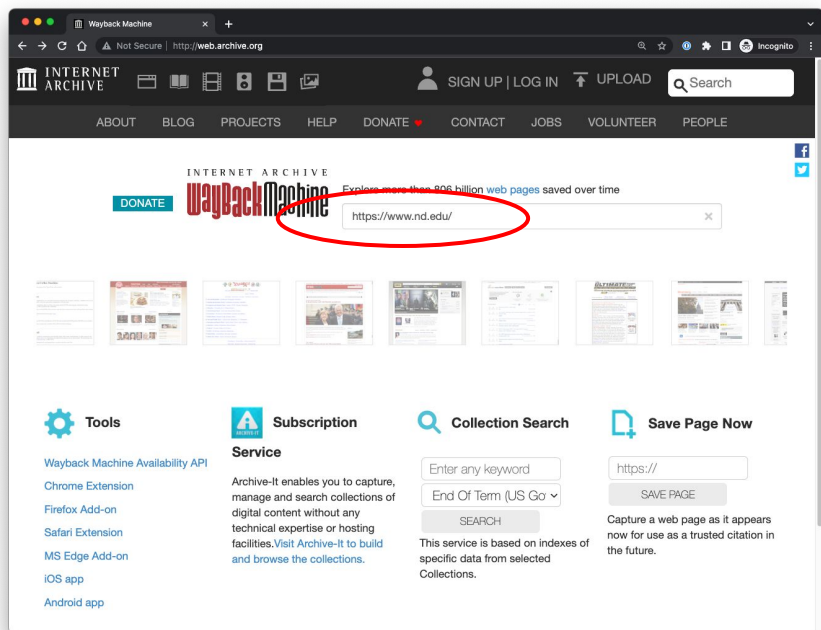


<https://twitter.com/IlyaKreymer/status/1288193177279541248>

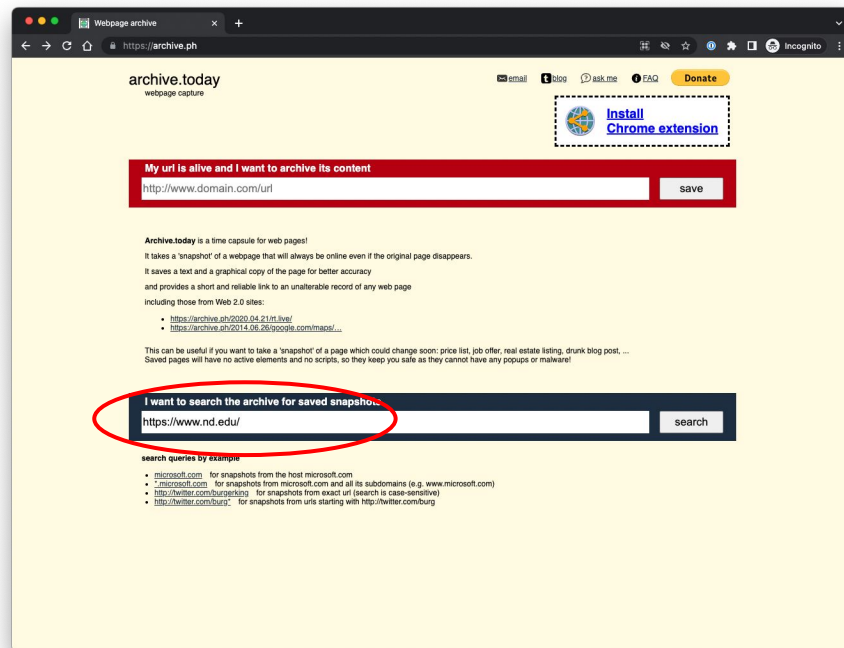
Mat Kelly and Michele C. Weigle, "WARCreate - Create Wayback-Consumable WARC Files from Any Webpage," In JCDL 2012, <https://www.cs.odu.edu/~mweigle/papers/kelly-jcdl12.pdf>

Query web archives for https://www.nd.edu/

<https://web.archive.org/>

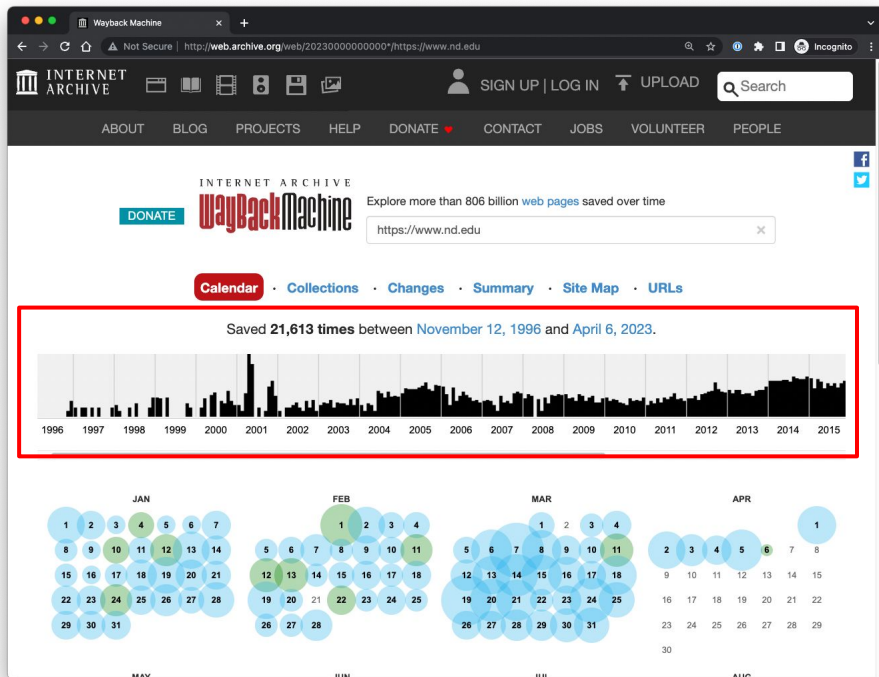


<https://archive.today/>

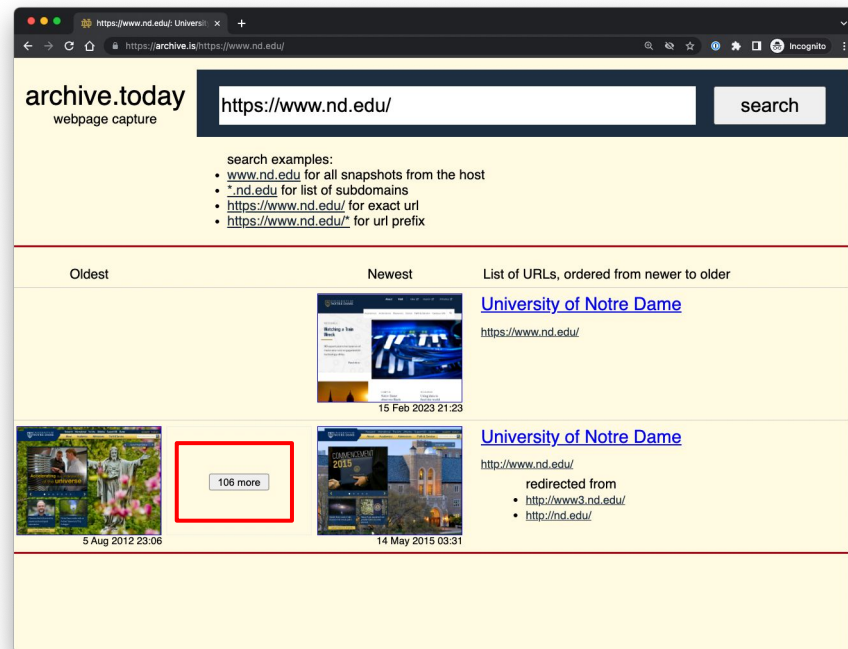


List of Mementos for <https://www.nd.edu/>

http://web.archive.org/web/20230000000000*/https://www.nd.edu



<https://archive.is/https://www.nd.edu/>



Navigating the Wayback Machine

<https://web.archive.org/web/19990208021711/https://www.nd.edu/>

The screenshot shows the Wayback Machine interface for the URL <https://www.nd.edu/>. The top navigation bar includes links for ABOUT, BLOG, PROJECTS, HELP, DONATE, CONTACT, JOBS, VOLUNTEER, and PEOPLE. Below the navigation bar, the Internet Archive logo is displayed, along with the text "Explore more than 806 billion web pages saved over time". A search bar is present, and a "Calendar" button is highlighted. The main content area shows a calendar for the year 1999, with a bar chart indicating the number of snapshots taken each day. A tooltip for February 6, 1999, at 02:17:11 is shown, indicating 8 snapshots were taken at that time. The URL bar shows the full path: <http://web.archive.org/web/19990208021711/https://www.nd.edu/>.

The screenshot shows the Notre Dame Email Home Page as it appeared in 1999. The page features the Notre Dame logo and the text "university of notre dame network engineering". The main heading is "Notre Dame Email Home Page". Below the heading, there are sections for "You can search:", "You can administer your account:", and "You can work with LISTSERV lists:". The "You can search:" section lists the "X.500 Directory Service" (Find email addresses of Notre Dame Students, Faculty, and Staff). The "You can administer your account:" section lists several options: "Remove Forwarding Of Your ND Email Account" (All mail is sent to your ND mail account. You will need to configure a mail client like Netscape to read your mail.), "Forward Your ND Email" (All mail is sent to the address you specify, ex. bob@aol.com. Do NOT use this to forward mail to yourself at nd.edu. Use the link directly above to receive mail at nd.edu), "Configure Vacation Auto-reply", and "View IMAP quota information". The "You can work with LISTSERV lists:" section lists the "ND LISTSERV Home Page". The "You can learn Eudora:" section lists the "Eudora Tutorial" (Windows PC). The URL bar shows the full path: <http://web.archive.org/web/19990208021711/https://www.nd.edu/>.

Navigating the Wayback Machine

<https://web.archive.org/web/20010118210700/https://www.nd.edu/>

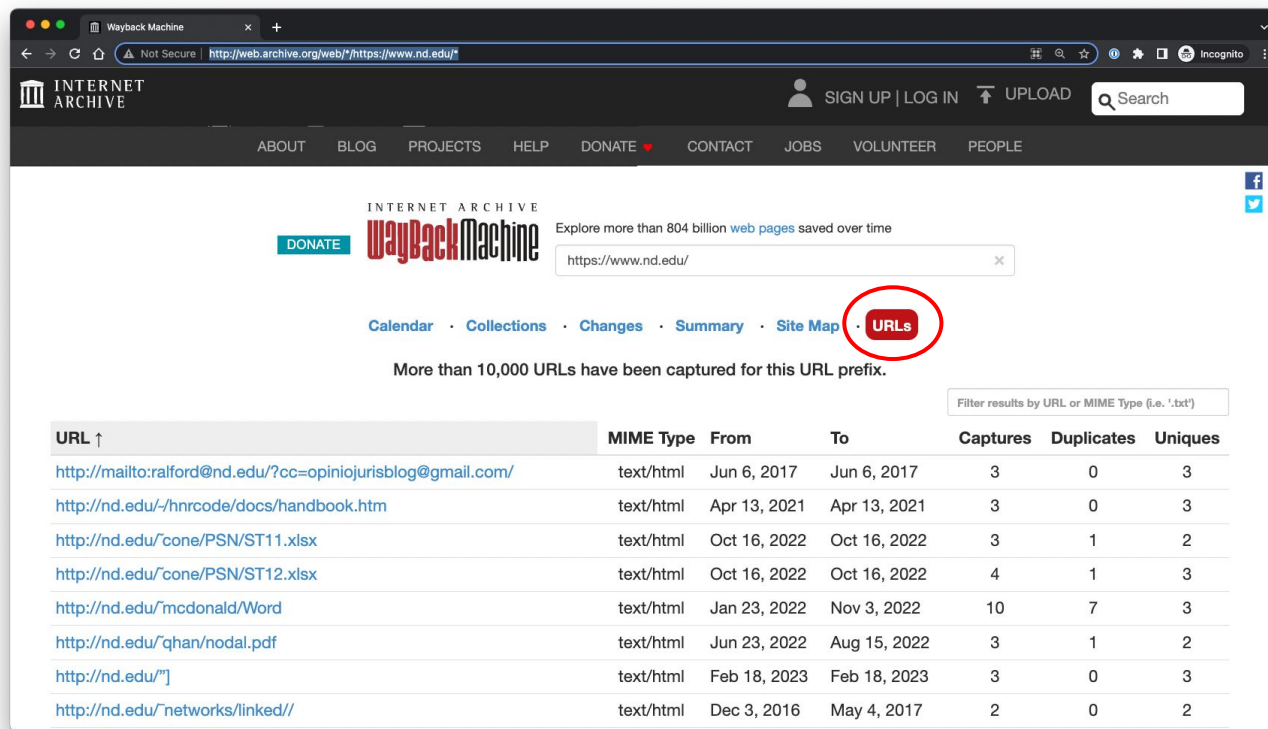
The screenshot shows the Wayback Machine interface. At the top, there's a search bar with the URL <https://www.nd.edu/>. Below the search bar, there's a navigation menu with links like ABOUT, BLOG, PROJECTS, HELP, DONATE, CONTACT, JOBS, VOLUNTEER, and PEOPLE. The main content area displays the Internet Archive logo and a message: "Explore more than 806 billion web pages saved over time". A calendar for the year 2001 is shown, with a specific date, January 18, 2001, highlighted. A tooltip for this date shows the time 21:07:00. The bottom of the interface shows the URL <https://web.archive.org/web/20010118210700/https://www.nd.edu/>.

The screenshot shows the University of Notre Dame website as it appeared in 2001. The header includes the University of Notre Dame logo and a navigation menu with links like TODAY@NOTRE DAME, DIRECTORIES, and CAMPUS MAP. The main content area features a large image of the Basilica of the Sacred Heart and a section titled "Get ready for the 2001 Internship Fair Jan. 24 at the Joyce". The footer contains contact information for the University of Notre Dame and a copyright notice for 2000.

Can also issue a prefix query to the Wayback Machine

http://web.archive.org/web/*/https://www.nd.edu/

Lists all URLs
archived that start
with www.nd.edu
or nd.edu



INTERNET ARCHIVE

WayBackMachine

Explore more than 804 billion web pages saved over time

https://www.nd.edu/

Calendar · Collections · Changes · Summary · Site Map · **URLs**

More than 10,000 URLs have been captured for this URL prefix.

Filter results by URL or MIME Type (i.e. '.txt')

URL ↑	MIME Type	From	To	Captures	Duplicates	Uniques
http://mailto:ralford@nd.edu/?cc=opiniojurisblog@gmail.com/	text/html	Jun 6, 2017	Jun 6, 2017	3	0	3
http://nd.edu/~hnrdoc/docs/handbook.htm	text/html	Apr 13, 2021	Apr 13, 2021	3	0	3
http://nd.edu/~cone/PSN/ST11.xlsx	text/html	Oct 16, 2022	Oct 16, 2022	3	1	2
http://nd.edu/~cone/PSN/ST12.xlsx	text/html	Oct 16, 2022	Oct 16, 2022	4	1	3
http://nd.edu/~mcdonald/Word	text/html	Jan 23, 2022	Nov 3, 2022	10	7	3
http://nd.edu/~qhan/nodal.pdf	text/html	Jun 23, 2022	Aug 15, 2022	3	1	2
http://nd.edu/"]	text/html	Feb 18, 2023	Feb 18, 2023	3	0	3
http://nd.edu/~networks/linked//	text/html	Dec 3, 2016	May 4, 2017	2	0	2

Querying Multiple Web Archives Using Memento Time Travel



The screenshot shows the Memento Time Travel interface. At the top, there's a search bar with the URL `http://www.nd.edu/` and a dropdown menu set to 'Time Travel'. Below the search bar, a timeline shows the requested date `25 Apr 2014 00:39:23 GMT`. The results section lists several Mementos from different web archives:

- Internet Archive Memento, 3 hours before** (Embed): `https://web.archive.org/web/20140425003923/http://www.nd.edu/` (25 Apr 2014 00:39:23 GMT [-3 hours from requested date])
- Previous Memento** (25 Apr 2014 00:23:44 GMT [-4 hours])
- Next Memento** (25 Apr 2014 10:04:52 GMT [+5 hours])
- First Memento** (12 Nov 1996 00:45:59 GMT [-17 years 168 days])
- Last Memento** (09 Apr 2023 23:05:33 GMT [+8 years 351 days])
- All captures from Internet Archive between 1996 and 2023**
- archive.is Memento, 4 days before** (Embed): `http://archive.md/20140421023111/http://www.nd.edu/` (21 Apr 2014 02:31:11 GMT [-4 days from requested date])
- Previous Memento** (21 Apr 2014 02:31:11 GMT [-4 days])
- Next Memento** (02 May 2014 02:33:00 GMT [+6 days])
- First Memento** (05 Aug 2012 23:06:10 GMT [-1 year 262 days])
- Last Memento** (15 Feb 2023 21:23:54 GMT [+8 years 298 days])
- All captures from archive.is between 2012 and 2023**
- Stanford Web Archive Memento, 17 days after** (Embed): `https://swa.stanford.edu/was/20140512070512mp_/http://www.nd.edu/` (12 May 2014 07:05:12 GMT [+17 days from requested date])
- Previous Memento** (data not provided)
- Next Memento** (data not provided)

On the right side of the interface, there are several promotional banners for Memento, including 'experience web time travel', 'install Memento for Chrome', 'enable web time travel' (with a Mediawiki logo), 'say no to "404 Not Found"', 'Robust Links', and 'Robustify your Zotero links!'.

<https://timetravel.mementoweb.org/>

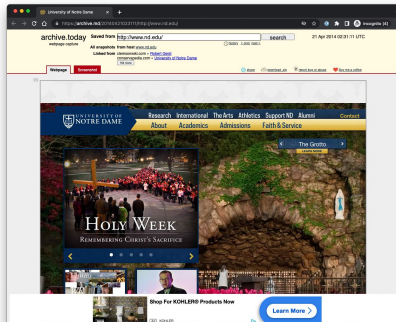
<http://timetravel.mementoweb.org/list/20140425043528/http://www.nd.edu/>

Herbert Van de Sompel, Michael L. Nelson, and Robert D. Sanderson, "HTTP Framework for Time-Based Access to Resource States -- Memento", RFC 7089, <https://www.rfc-editor.org/rfc/rfc7089.html>

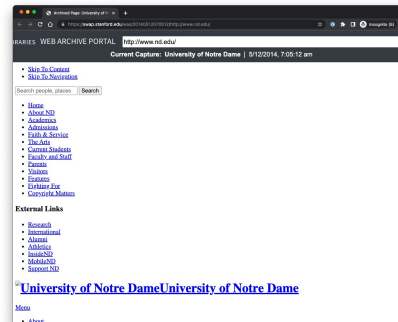
Closest Mementos to April 25, 2014



Internet Archive - **20140425003923**



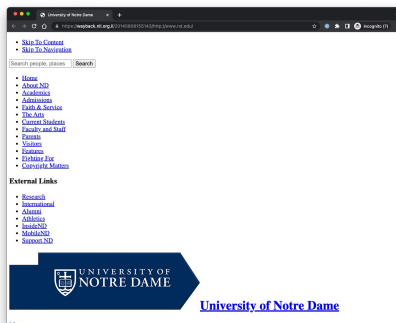
archive.is - **20140421023111**



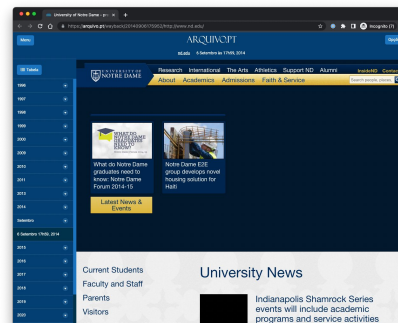
Stanford - **20140512070512**



Library of Congress - **20140307142751**



National Library of Israel - **20140806155143**



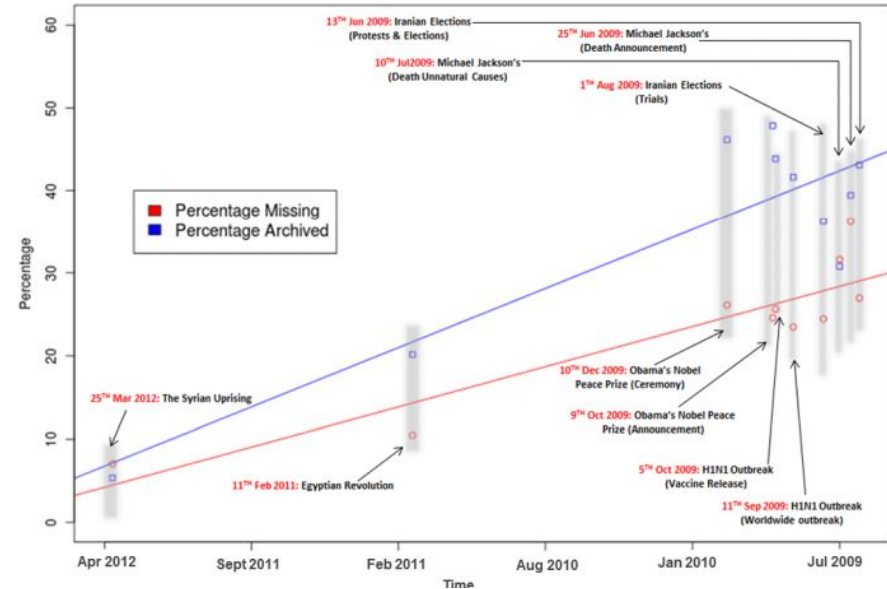
Arquivo.pt - **20140906175952**

Using Web Archives to Document Social Movements and Disinformation - Practice, Ethics, and Challenges

Using Web Archives to Document Social Movements - **Practice**

Egyptian Revolution (2011)

- Called the "Facebook Revolution" because of the large role that social media played in helping protesters to organize
- Study revealed that after 1 year, 10% of the resources *linked to* in tweets were no longer available
 - only 20% of the resources had been archived
- Percentage lost and percentage archived increased over time
- This held for multiple events between June 2009 - March 2012



Hany SalahEldeen, Losing My Revolution: A year after the Egyptian Revolution, 10% of the social media documentation is gone.,

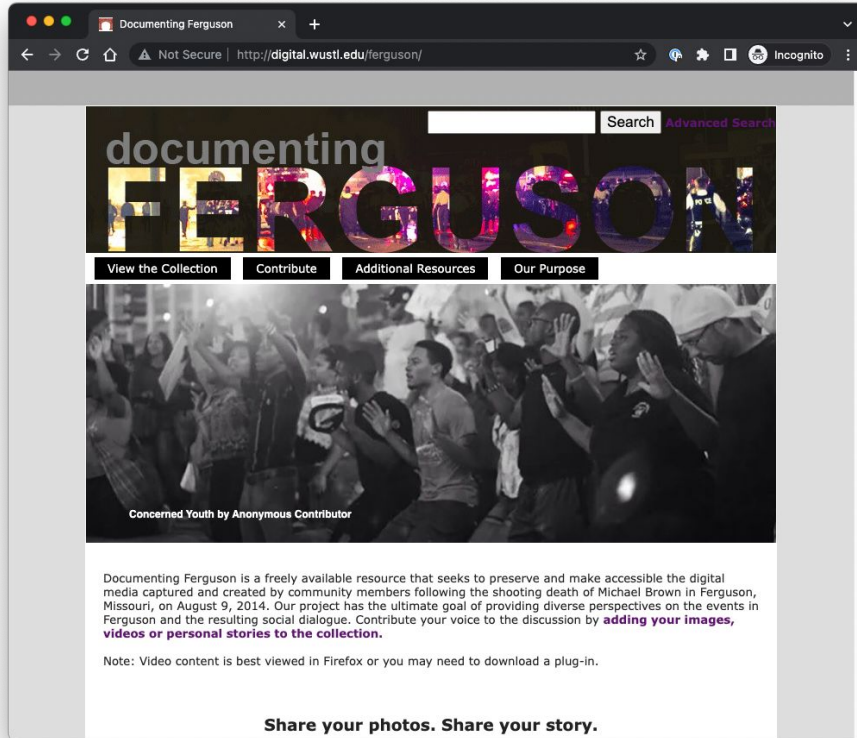
<https://ws-dl.blogspot.com/2012/02/2012-02-11-losing-my-revolution-year.html>

Hany SalahEldeen and Michael L. Nelson, Losing My Revolution: How Many Resources Shared on Social Media Have Been Lost?, TPDL 2012, <https://arxiv.org/abs/1209.3026>

Hany SalahEldeen and Michael L. Nelson, Resurrecting My Revolution: Using Social Link Neighborhood in Bringing Context to the Disappearing Web, TPDL 2013,

<https://arxiv.org/abs/1309.2648>

Documenting Ferguson (2014)



<http://digital.wustl.edu/ferguson/>



<https://twitter.com/deray/status/549442824752889856>

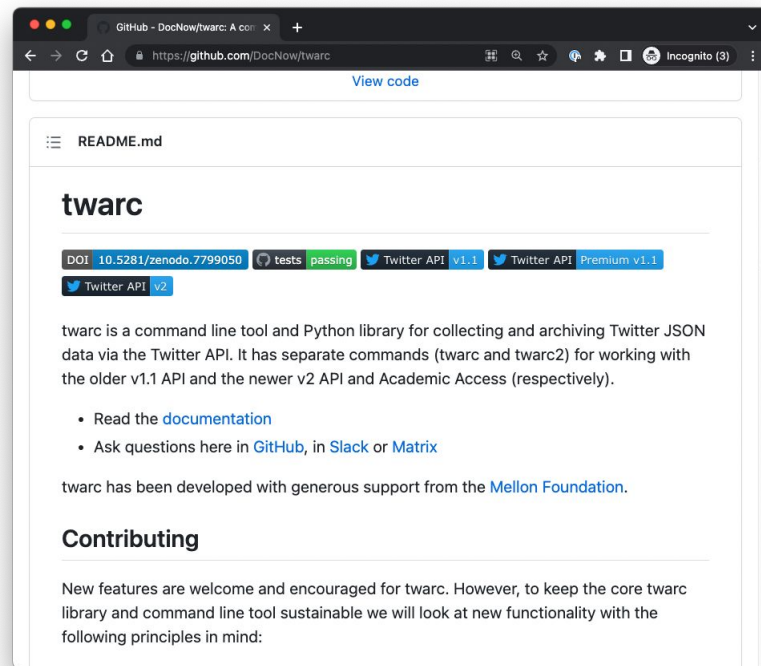
Collecting Large Tweet Datasets

- Created via the Twitter API, using tools such as twarc
- Download JSON content of tweets, including tweet text and metadata

```
1 {
2   "data": [
3     {
4       "id": "1460323737035677698",
5       "text": "Introducing a new era for the Twitter Developer Platform! nn👉The Twitter AP
6       "edit_history_tweet_ids": [
7         "1460323737035677698"
8       ]
9     },
10    {
11      "id": "1519781379172495360",
12      "text": "Our mission remains just as important as ever: to deliver an open platform t
13      "edit_history_tweet_ids": [
14        "1519781379172495360"
15      ]
16    }
17  ]
18 }
```

<https://developer.twitter.com/en/docs/twitter-api/tweets/lookup/api-reference/get-tweets#tab0>

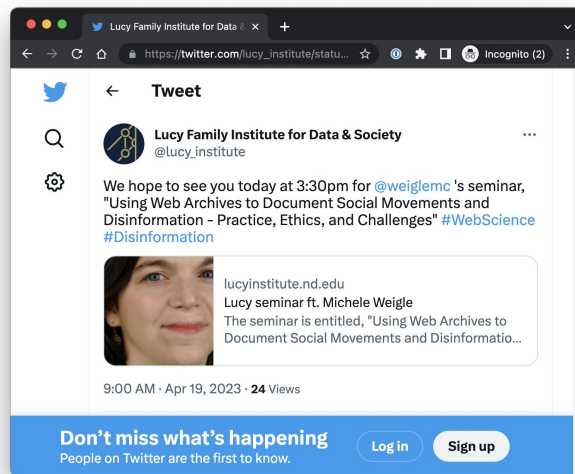
<https://github.com/DocNow/twarc>



Twitter's TOS only allows sharing large datasets as Tweet IDs

1	723539060208234498
2	723539066491383812
3	723539091988578304
4	723539105905315840
5	723539117527568384
6	723539120207876096
7	723539139686195200
8	723539151342194688
9	723539182400974849
10	723539188683931648
11	723539198721036292
12	723539237165887488
13	723539269567025152


https://twitter.com/lucy_institute/status/1648672988525387777
tweet id



Turning tweet ID into tweet content (JSON format) requires "hydration" using Twitter's API

Ferguson Twitter Archive

<https://catalog.docnow.io/datasets/20161224-ferguson-tweets/>

 **Catalog**

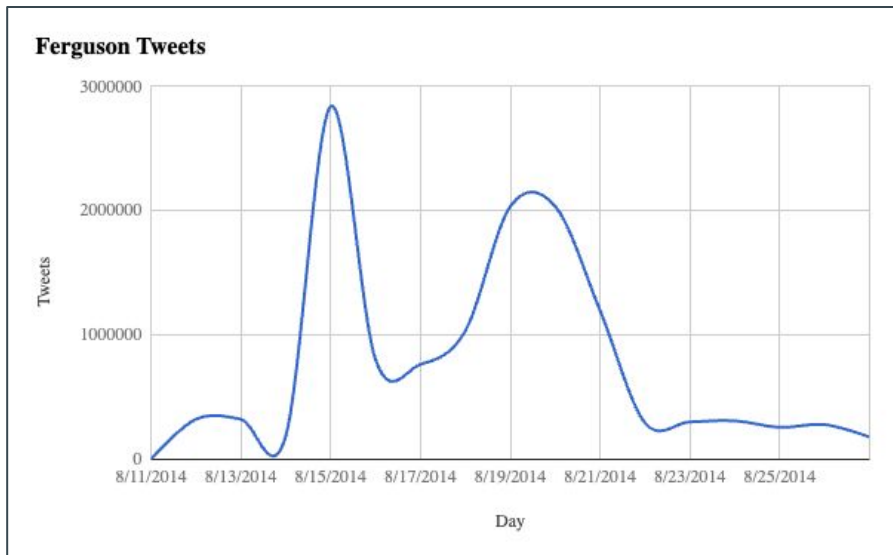
Title: Ferguson Tweets	Description: This item represents a collection of 13,480,000 tweet IDs that mentioned 'ferguson' from 2014-08-10 to 2014-08-27 and 15,080,078 tweet IDs that mention 'ferguson' between 2014-11-11 and 2014-12-08. The first set includes tweets for the two week period after the shooting of Michael Brown, and the second range includes tweets around the grand jury's decision not to indict police officer Darren Wilson which was announced on 2014-11-24. The first set of tweets were collected by Ed Summers at the University of Maryland and the second was a collaboration between Molly Loyd, Gregory Coleman, Kimberly Lamke, Benjamin Sugar and Ed Summers.
Repository: Internet Archive	
Repository URL: https://archive.org/details/ferguson-tweet-ids	
Creator(s): Ed Summers	
Subjects: Activism Blacklivesmatter	
Dates: 08/10/2014 - 08/27/2014 11/11/2014 - 12/08/2014	
Number of Tweets: 28,560,078	

28M tweet IDs

"ferguson"

2014-08-10 - 2014-08-27

2014-11-11 - 2014-12-08



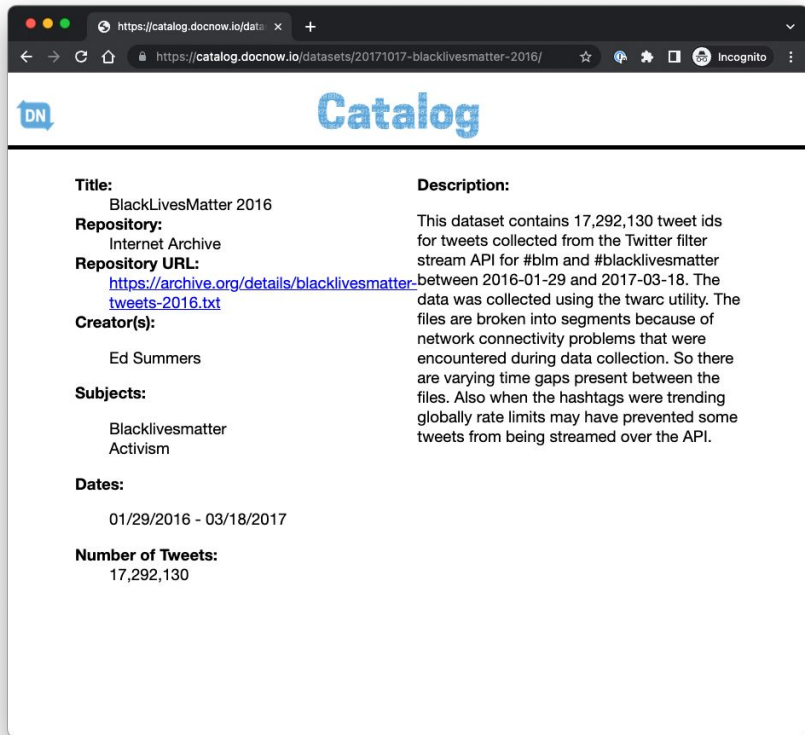
Ed Summers, A Ferguson Twitter Archive,

<https://inkdroid.org/2014/08/30/a-ferguson-twitter-archive/>

Archiving #BlackLivesMatter (2016)

<https://catalog.docnow.io/datasets/20171017-blacklivesmatter-2016/>

17M tweet IDs
#blm, #blacklivesmatter
2016-01-29 - 2017-03-18



The screenshot shows a web browser window with the URL <https://catalog.docnow.io/datasets/20171017-blacklivesmatter-2016/>. The page features a 'Catalog' header and a detailed description of the dataset. The description states that the dataset contains 17,292,130 tweet IDs for tweets collected from the Twitter filter stream API for #blm and #blacklivesmatter between 2016-01-29 and 2017-03-18. The files are broken into segments due to network connectivity problems. The dataset was collected using the twarc utility. The files are broken into segments because of network connectivity problems that were encountered during data collection. So there are varying time gaps present between the files. Also when the hashtags were trending globally rate limits may have prevented some tweets from being streamed over the API.

Title: BlackLivesMatter 2016

Repository: Internet Archive

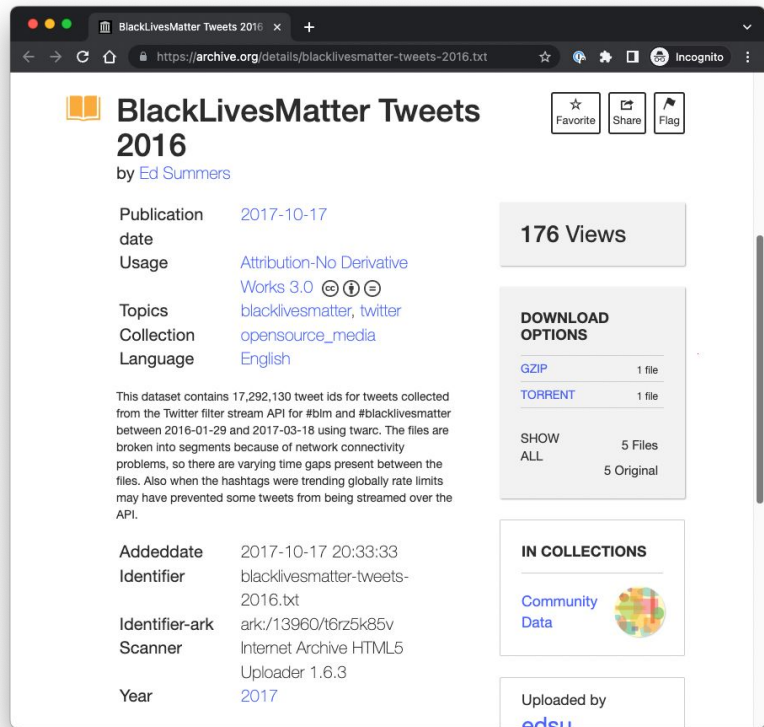
Repository URL: <https://archive.org/details/blacklivesmatter-tweets-2016.txt>

Creator(s): Ed Summers

Subjects: Blacklivesmatter, Activism

Dates: 01/29/2016 - 03/18/2017

Number of Tweets: 17,292,130



The screenshot shows a web browser window with the URL <https://archive.org/details/blacklivesmatter-tweets-2016.txt>. The page displays the title 'BlackLivesMatter Tweets 2016' by Ed Summers. It includes a table with metadata such as Publication date (2017-10-17), Usage (Attribution-No Derivative Works 3.0), Topics (blacklivesmatter, twitter), Collection (opensource_media), and Language (English). The page also shows a description of the dataset, download options (GZIP and TORRENT), and a section for 'IN COLLECTIONS'.

BlackLivesMatter Tweets 2016
by Ed Summers

Publication date: 2017-10-17

Usage: Attribution-No Derivative Works 3.0

Topics: blacklivesmatter, twitter

Collection: opensource_media

Language: English

This dataset contains 17,292,130 tweet ids for tweets collected from the Twitter filter stream API for #blm and #blacklivesmatter between 2016-01-29 and 2017-03-18 using twarc. The files are broken into segments because of network connectivity problems, so there are varying time gaps present between the files. Also when the hashtags were trending globally rate limits may have prevented some tweets from being streamed over the API.

Addeddate: 2017-10-17 20:33:33

Identifier: blacklivesmatter-tweets-2016.txt

Identifier-ark: ark:/13960/t6rz5k85v

Scanner: Internet Archive HTML5 Uploader 1.6.3

Year: 2017

176 Views

DOWNLOAD OPTIONS

GZIP 1 file

TORRENT 1 file

SHOW ALL 5 Files 5 Original

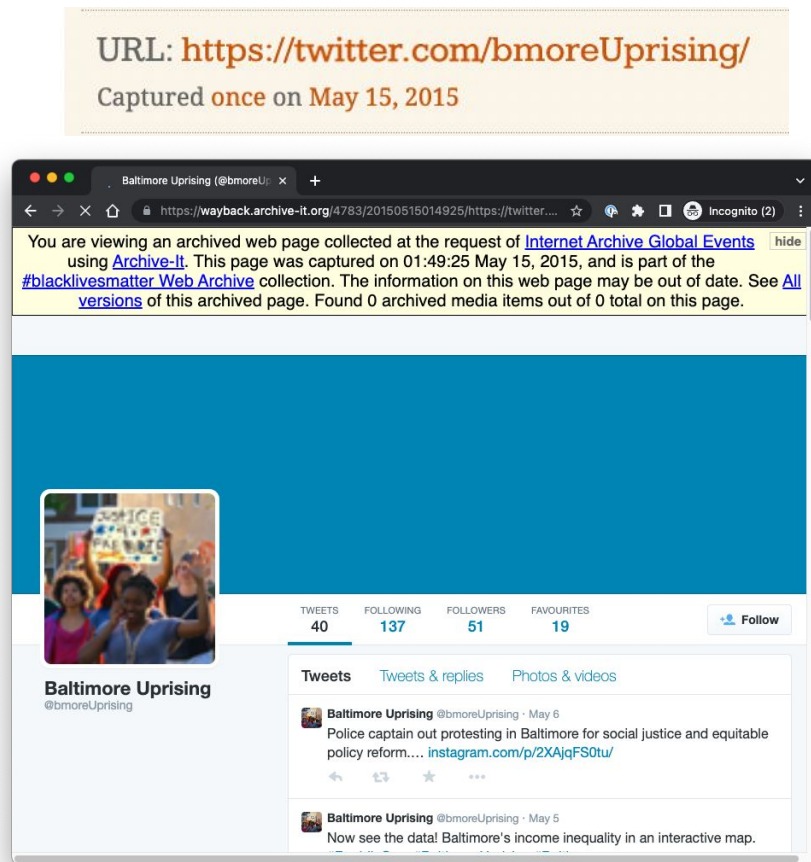
IN COLLECTIONS

Community Data

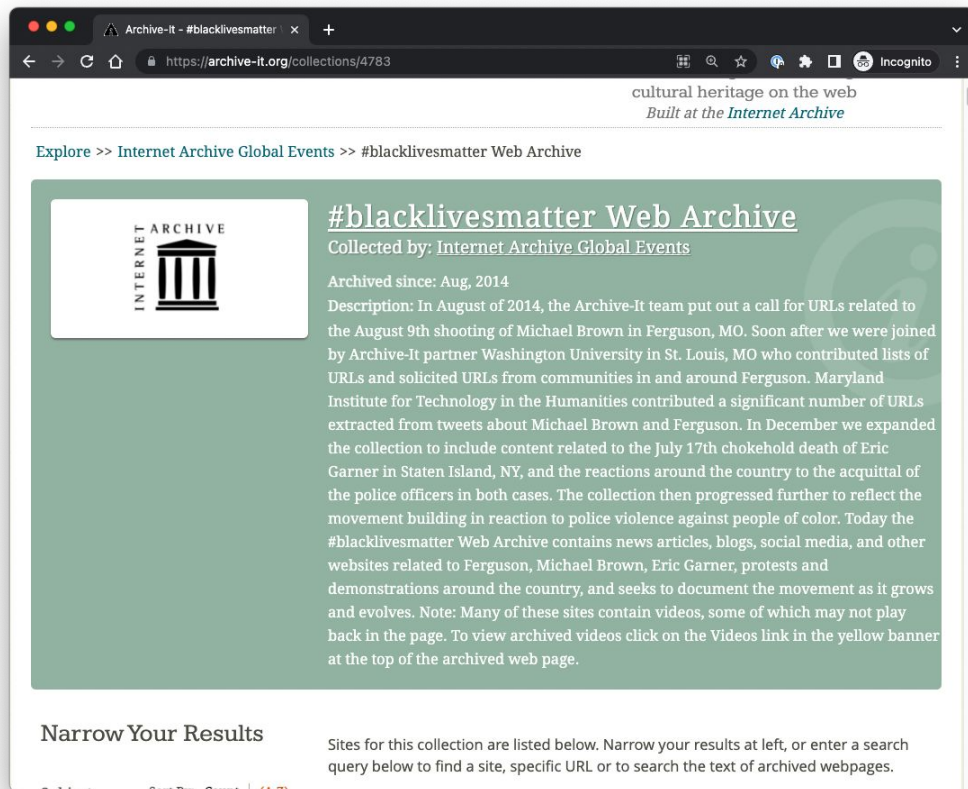
Uploaded by edu

Web Archiving Tweets

- Web archives access tweets via the web (not Twitter's API)
- Can archive a single tweet (using the post URL) or a set of tweets from an account (using the account URL)
- Does not depend upon the Twitter API for replay
- Can view deleted tweets and tweets from deleted/suspended accounts
- Processing tweet content requires web scraping
- Archiving and later analysis occurs on a much smaller scale



#blacklivesmatter Web Archive



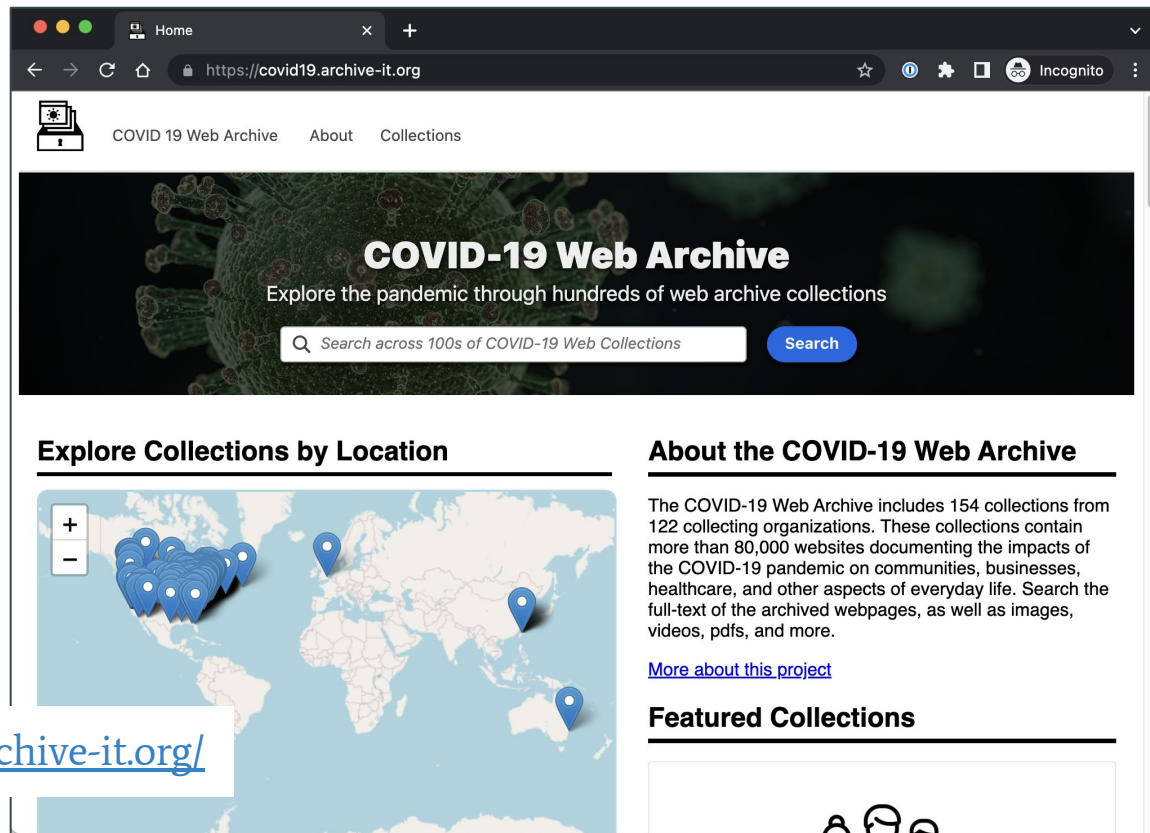
<https://archive-it.org/collections/4783>

1060 URLs, including 95
Twitter accounts, hashtags,
and search results

started Aug 2014

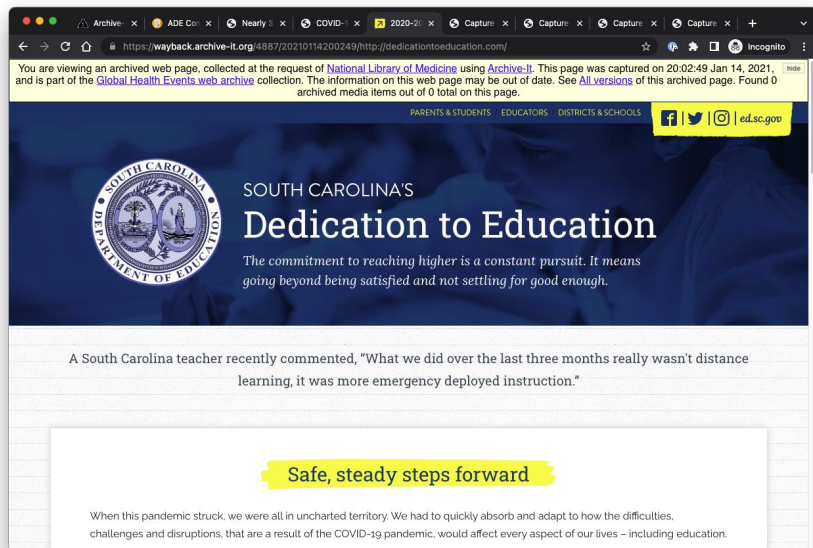
COVID-19 (2020)

154 collections at Archive-It

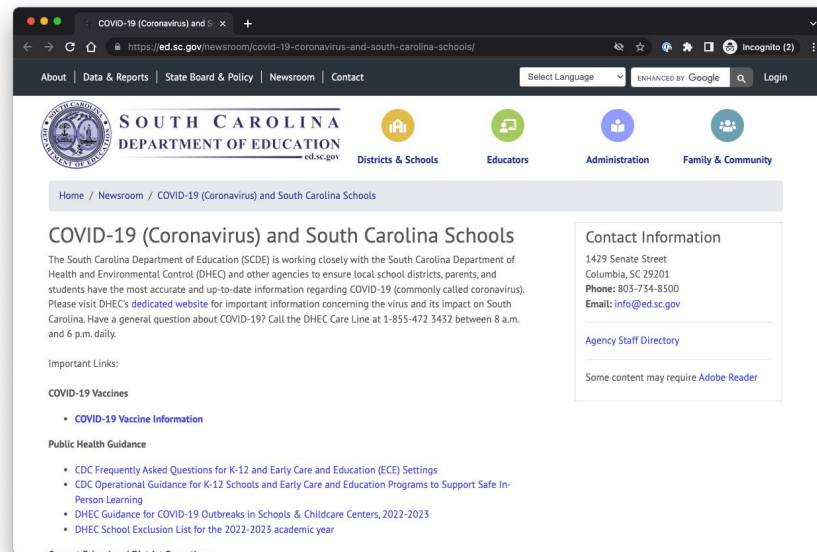


<https://covid19.archive-it.org/>

South Carolina Dept of Education, Dedication to Education Page



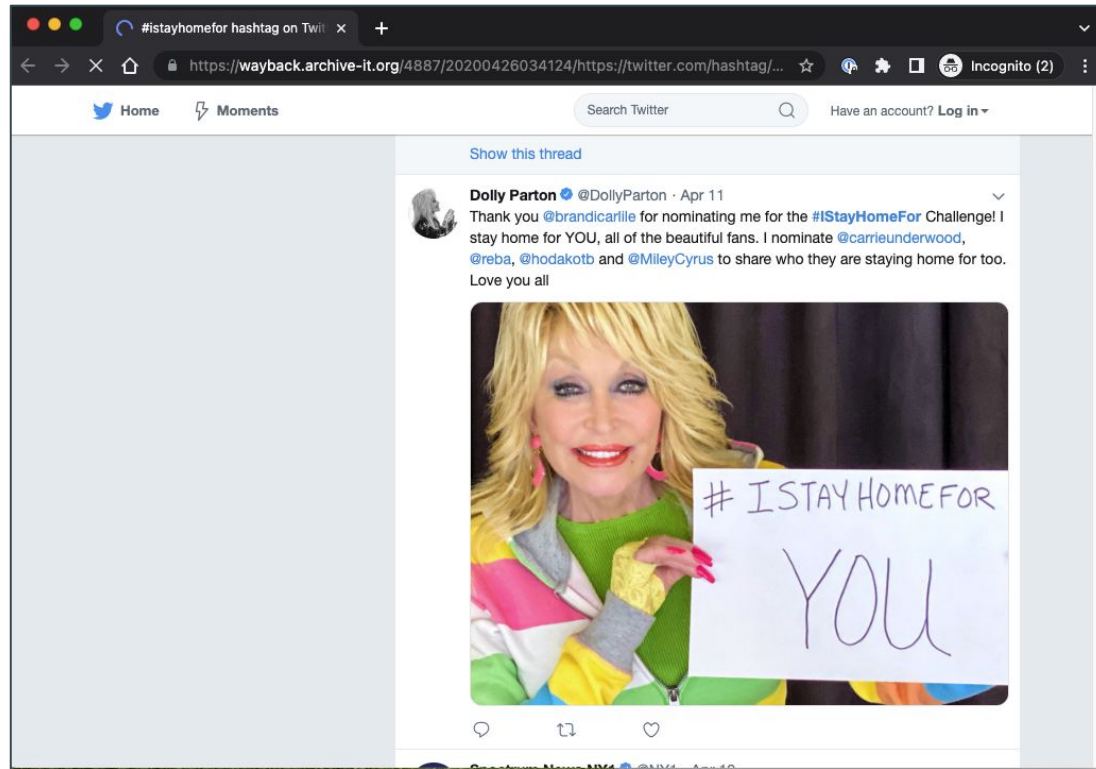
January 14, 2021



today
redirects to SC DOE's COVID Newsroom,
<https://ed.sc.gov/newsroom/covid-19-coronavirus-and-south-carolina-schools/>

https://wayback.archive-it.org/4887/*/http://dedicationtoeducation.com/

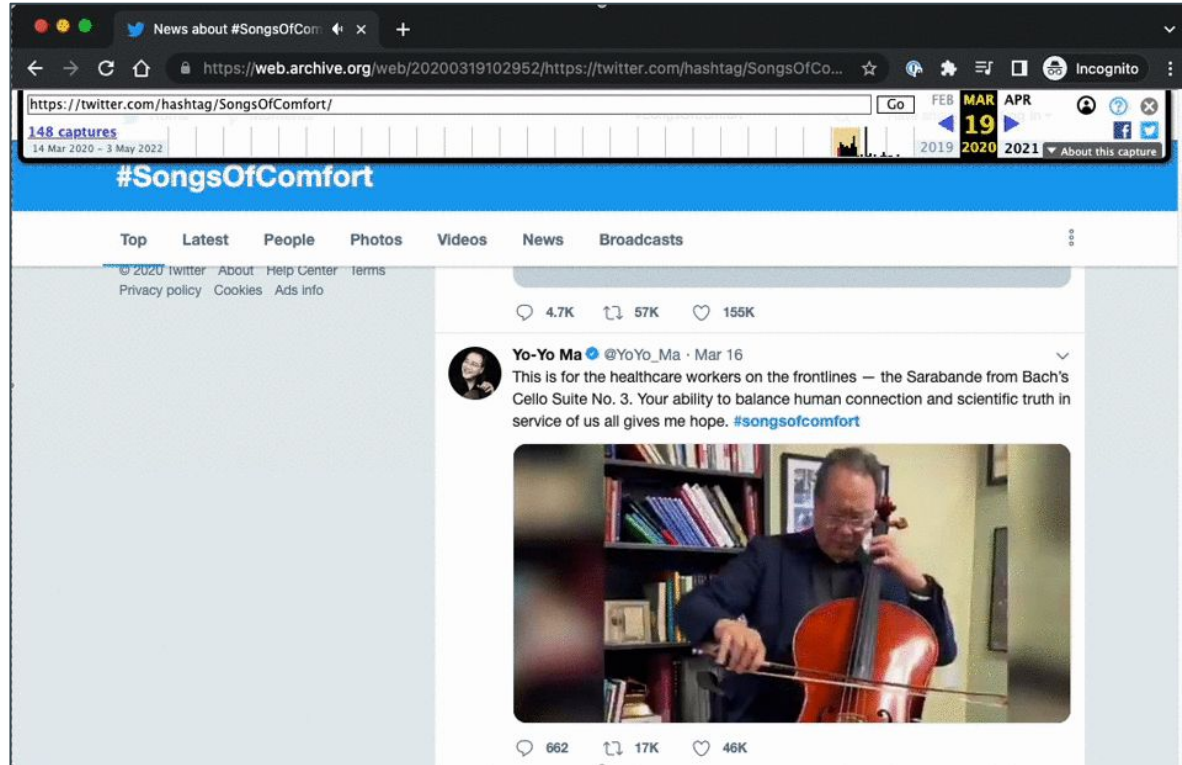
#IStayHomeFor on Twitter (April 2020)



April 26, 2020

<https://wayback.archive-it.org/4887/20200426034124/https://twitter.com/hashtag/IStayHomeFor/>

#SongsOfComfort on Twitter (March 2020)



March 19, 2020

<https://web.archive.org/web/20200319102952/https://twitter.com/hashtag/SongsOfComfort/>

Using Web Archives to Document Social Movements and Disinformation - Practice, Ethics, and Challenges

Using Web Archives to Document Disinformation - **Practice**

Journalists have been using web archives for a while now

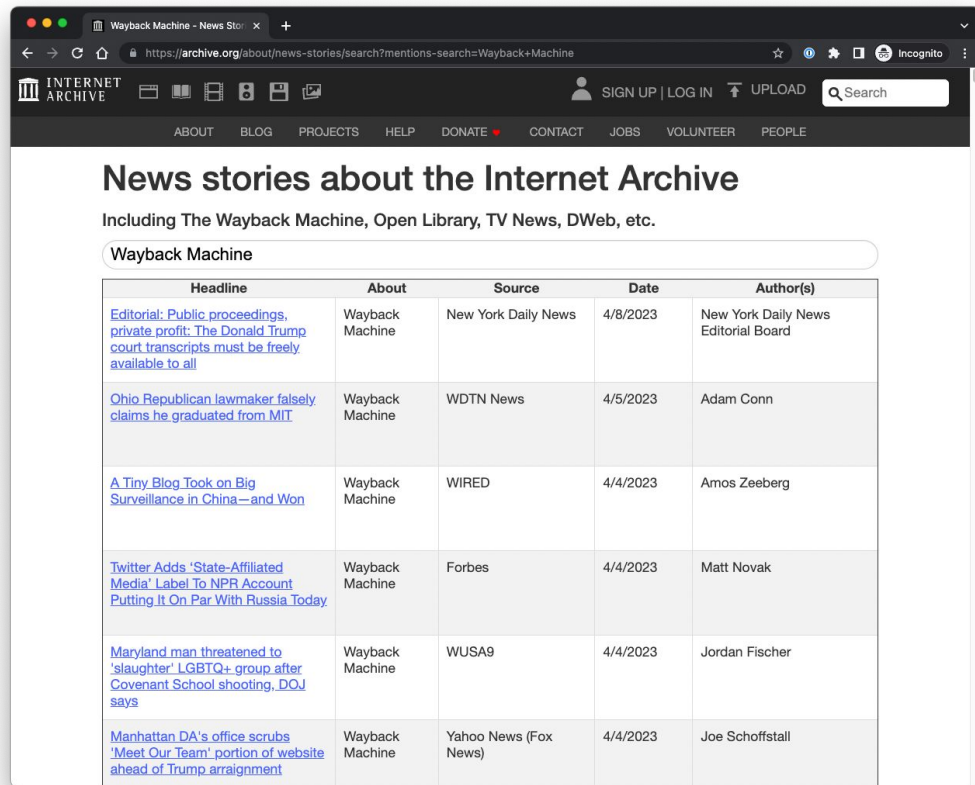


"God bless you, Wayback Machine"
- Rachel Maddow, Dec 16, 2016



Last Week Tonight, Mar 18, 2018

Internet Archive even tracks its mentions in news stories



Wayback Machine - News Stories

https://archive.org/about/news-stories/search?mentions-search=Wayback+Machine

INTERNET ARCHIVE

SIGN UP | LOG IN | UPLOAD

ABOUT BLOG PROJECTS HELP DONATE CONTACT JOBS VOLUNTEER PEOPLE

News stories about the Internet Archive

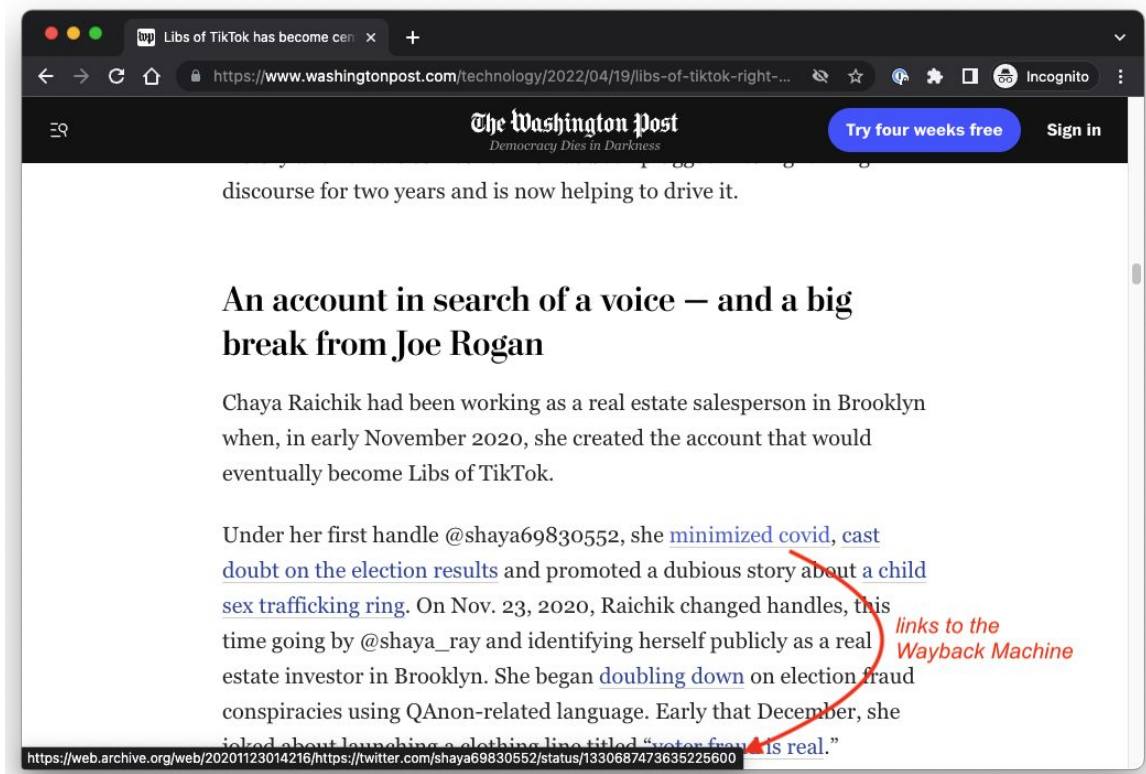
Including The Wayback Machine, Open Library, TV News, DWeb, etc.

Wayback Machine

Headline	About	Source	Date	Author(s)
Editorial: Public proceedings, private profit: The Donald Trump court transcripts must be freely available to all	Wayback Machine	New York Daily News	4/8/2023	New York Daily News Editorial Board
Ohio Republican lawmaker falsely claims he graduated from MIT	Wayback Machine	WDTN News	4/5/2023	Adam Conn
A Tiny Blog Took on Big Surveillance in China—and Won	Wayback Machine	WIRED	4/4/2023	Amos Zeeberg
Twitter Adds 'State-Affiliated Media' Label To NPR Account Putting It On Par With Russia Today	Wayback Machine	Forbes	4/4/2023	Matt Novak
Maryland man threatened to 'slaughter' LGBTQ+ group after Covenant School shooting, DOJ says	Wayback Machine	WUSA9	4/4/2023	Jordan Fischer
Manhattan DA's office scrubs 'Meet Our Team' portion of website ahead of Trump arraignment	Wayback Machine	Yahoo News (Fox News)	4/4/2023	Joe Schoffstall

<https://archive.org/about/news-stories/search?mentions-search=Wayback+Machine>

Example of news story with links to the Wayback Machine



Taylor Lorenz, "Meet the woman behind Libs of TikTok", *Washington Post*, April 2022,
<https://www.washingtonpost.com/technology/2022/04/19/lib-...>

Researchers submit webpages to Internet Archive for later reference

Election Integrity Partnership archived posts about the 2020 Election at Internet Archive so they could go back and write about them later.

The long fuse: Misinformation and the 2020 election

THE LONG FUSE: MISINFORMATION AND THE 2020 ELECTION

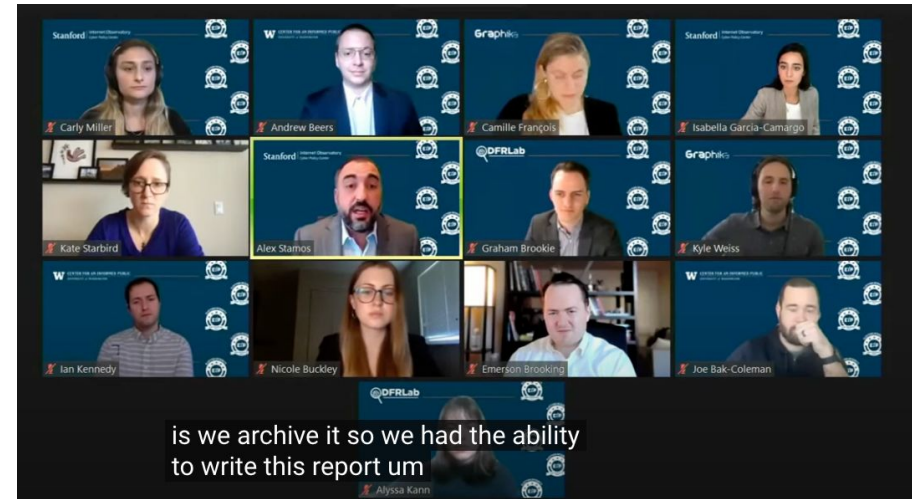
MARCH 3, 2021 | 3:00 - 4:30 P.M. EST

FIRESIDE CHAT: CHRIS KREBS,
FORMER DIRECTOR OF DHS-CISA

PRESENTATION: EXPERT BREAKDOWN OF THE
ELECTION INTEGRITY
PARTNERSHIP'S REPORT

Logos at the bottom: Election Integrity Partnership, DFRLab, Graphika, Stanford, Internet Observatory Cyber Policy Center, W CENTER FOR AN INFORMED PUBLIC UNIVERSITY of WASHINGTON

Watch on YouTube



Election Integrity Partnership, The long fuse: Misinformation and the 2020 election, <https://www.atlanticcouncil.org/event/the-long-fuse-eip-report/>, <https://www.youtube.com/watch?v=uKkaZI-EiiQ&t=4524s>

Unfortunately, web archives can be used to spread removed content



Mohandas Pai ✓
@TVMohandasPai

Covid-19 had us all fooled, but now we might have finally found its secret. Interesting! Any comments @drballalmanipal @satyamysore



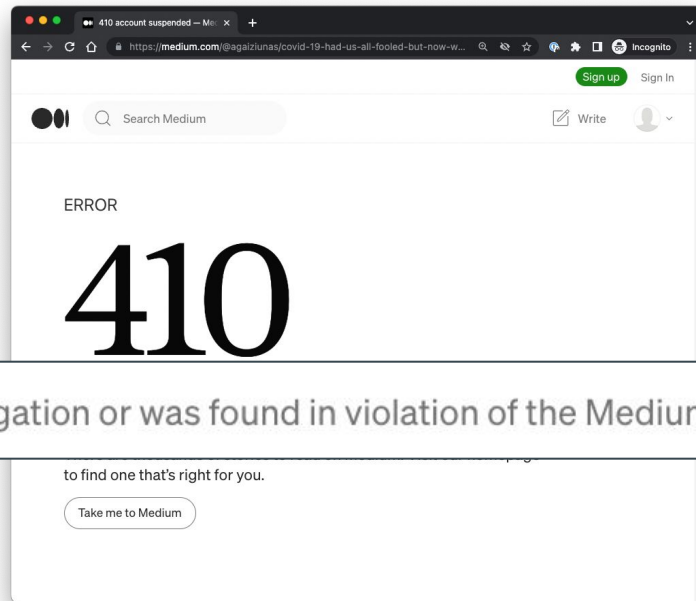
web.archive.org

Covid-19 had us all fo

In the last 3-5 days, a mountain of anecdotal evidence has come out of NYC, Italy, Spain, etc. ...

5:04 AM · Apr 8, 2020

<https://twitter.com/TVMohandasPai/status/1247812510373122048>

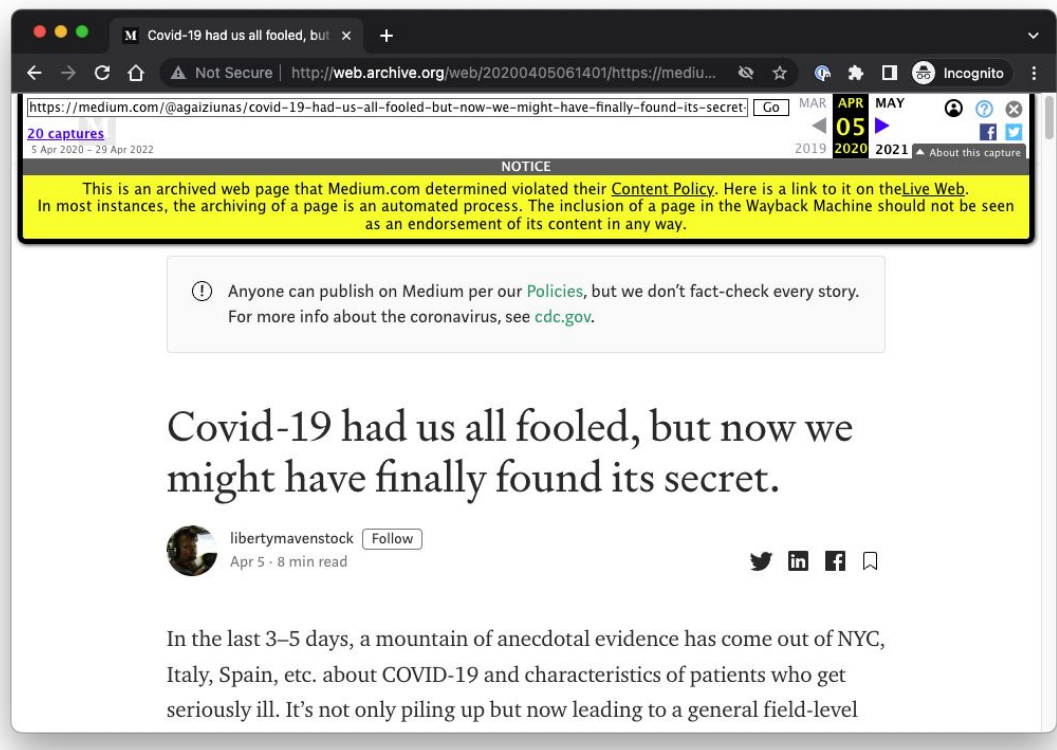


<https://medium.com/@agaiziunas/covid-19-had-us-all-fooled-but-now-we-might-have-finally-found-its-secret-91182386efcb>

Joan Donovan, "Covid hoaxes are using a loophole to stay alive—even after content is deleted", April 2020,

<https://www.technologyreview.com/2020/04/30/1000881/covid-hoaxes-zombie-content-wayback-machine-disinformation/>

However, Internet Archive is adding labels to some of this content



Mark Graham, "Fact Checks and Context for Wayback Machine Pages", Oct 2020, <https://blog.archive.org/2020/10/30/fact-checks-and-context-for-wayback-machine-pages/>

We need web archives to study original materials later, even if they contain misinformation

2022-09-28: Using Web Archives in Disinformation Research

By Michele Weigle - September 28, 2022

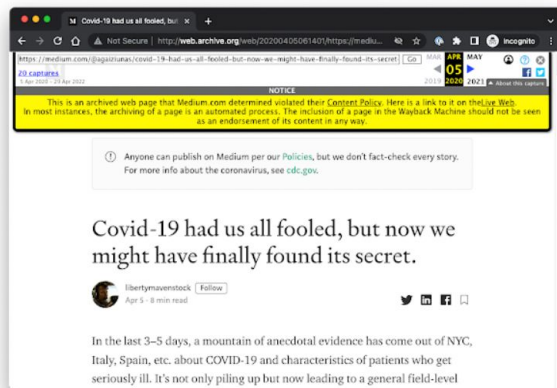


Figure: Example of content label added by the Internet Archive

Following on [Lesley Frew's post looking at how journalists use web archives](#), I wanted to highlight some of the ways that web archives have been used to study disinformation over the past few years and to bring together some of the work that our [Web Science and Digital Libraries \(WS-DL\) research group](#) has done in this area. My interest in the intersection of web archives and disinformation has largely stemmed from developing a [Disinformation lecture](#) for our [CS 432/532 Web](#)

“It’s not about trying to archive the stuff that’s true, but archive the conversation. All of that is what people are experiencing.”
-Mark Graham, Director of the Wayback Machine

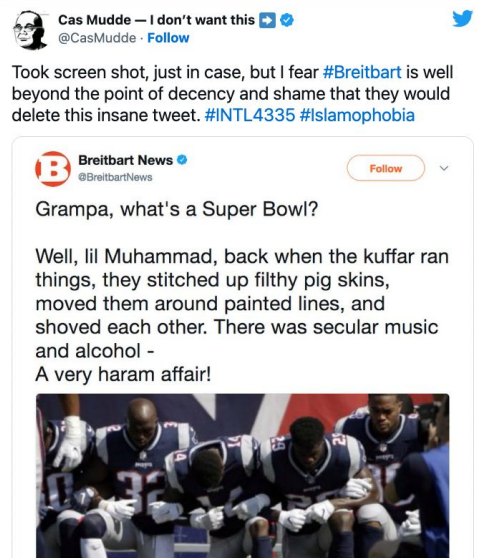
<https://www.ft.com/content/5be1f2ee-d60b-11e9-a0bd-ab8ec6435630>

Michele Weigle, "Using Web Archives in Disinformation Research", <https://ws-dl.blogspot.com/2022/09/2022-09-28-using-web-archives-in.html>

Using Web Archives to Find Deleted Tweets

2018-04-23: "Grampa, what's a deleted tweet?"

By Nauman Siddique - April 23, 2018



2019-04-17: Russell Westbrook, Shane Keisel, Fake Twitter Accounts, and Web Archives

By [Michael L. Nelson](#) - April 17, 2019



On March 11, 2019 in the [NBA](#), the Utah Jazz hosted their [Northwest Division](#) rivals, the Oklahoma City Thunder. During the game, a Utah fan (Shane Keisel) and a Oklahoma City player ([Russell Westbrook](#)) engaged in a verbal exchange, with the player stating the fan was directing racist comments to him and the fan admitting to heckling but denying that his comments were racist. The event was well documented (see, for example, this [Bleacher Report article](#)), and the following day the fan received a [lifetime ban](#) from all events at the [Vivint Smart Home Arena](#) and the player received a [\\$25k fine](#) from the NBA.

Disclaimer: I have no knowledge of what the fan said during the game, nor do I have an opinion regarding the appropriateness of the respective penalties. My interest is that after the game, the fan gave at least one [interview](#) with a TV station reporter in which he exposed his identity. That set off a rapidly evolving series of events with both real and fake Twitter accounts, which we unravel with the aid of multiple web archives. The [initial analysis](#) was performed by [Justin Whitlock](#) as a project in my CS 895 "[Web Archiving Forensics](#)" class; prior to Justin proposing it as a project topic, my only knowledge of this event was via the [Daily Show](#).

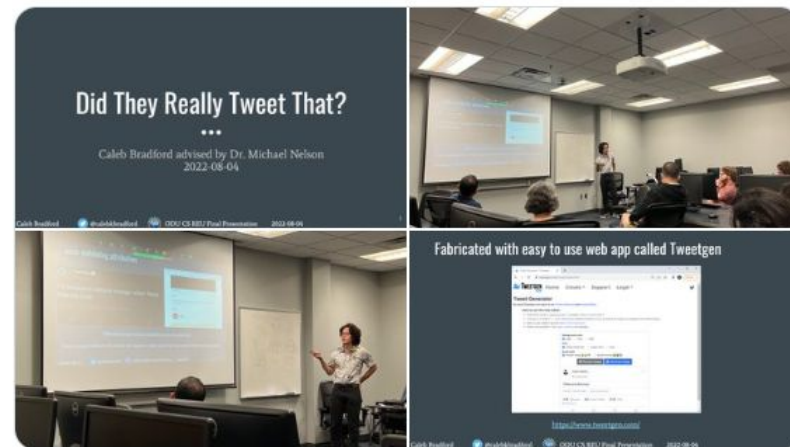
Michael Nelson, "Russell Westbrook, Shane Keisel, Fake Twitter Accounts, and Web Archives", <https://ws-dl.blogspot.com/2019/04/2019-04-17-russell-westbrook-shane.html>

Nauman Siddique, "Grampa, what's a deleted tweet?", Apr 2018, <https://ws-dl.blogspot.com/2018/04/2018-04-23-grampa-whats-deleted-tweet.html>

"Did They Really Tweet That?"



Final presentation @oducsreu presentations: @calebkbbradford presenting "Did They Really Tweet That?" (mentor: @phonedude mln). The project attempts to identify fabricated tweets through fact-checking and verification techniques. #oducsreu @WebSciDL @NirdsLab @oducs @ODUSCI @NSF



3:56 PM · Aug 4, 2022

<https://twitter.com/mahanama94/status/1555281601654755328>

Caleb Bradford and Michael L. Nelson, Did They Really Tweet That? Querying Fact-Checking Sites and Politwoops to Determine Tweet Misattribution, <https://arxiv.org/abs/2211.09681>

Tarannum Zaki, Disinformation Spread on Social Media through Screenshot Sharing: Dataset Description, <https://ws-dl.blogspot.com/2022/12/2022-12-12-disinformation-spread-on.html>

Find All Archived Tweets for a User at Internet Archive

http://web.archive.org/web/*/https://www.twitter.com/weiglemc/status/*

The screenshot shows the Internet Archive Wayback Machine interface. The search bar contains the URL `https://www.twitter.com/weiglemc/status/`. Below the search bar, it states "1,629 URLs have been captured for this URL prefix." A table lists the captured URLs with their MIME types and capture dates.

URL ↑	MIME Type	From	To	Captures
http://www.twitter.com/weiglemc/status/849056515227078656	unk	Dec 8, 2022	Dec 8, 2022	3
http://www.twitter.com/weiglemc/status/849056542628380679	unk	Dec 8, 2022	Dec 10, 2022	3
http://www.twitter.com/weiglemc/status/849209209312948226	unk	Aug 1, 2021	Dec 8, 2022	4
https://twitter.com/weiglemc/status/1003748822273679362/photo/1	text/html	Nov 18, 2022	Nov 18, 2022	1
https://twitter.com/weiglemc/status/1003748822273679362/photo/1?prefetchTimestamp=1668801279723	text/html	Nov 18, 2022	Nov 18, 2022	1

<http://web.archive.org/web/20211123000955/https://twitter.com/weiglemc/status/1003997508572471296>

The screenshot shows a tweet by Michele Weigle (@weiglemc) dated November 23, 2021. The tweet text is: "Today's opening session intro at #jcdl2018 included some shoutouts to @WebSciDL. Gotta submit a lot to get accepted. :)". The tweet includes a photo of a presentation slide titled "Best Paper Nominee Sessions" which lists several sessions and their speakers. The slide content is as follows:

- Session 5A: Alonso, Kandyas and Tremblay, How it Happened: Discovering and Archiving the Evolution of Social Signals
- Session 5A: Sanikawa, Jatowt and During, Digital History Meets Microblogging: Analyzing Collective Metatwitter
- Session 5C: Kelly, Nelson and Weigle, A Framework for Aggregating Private and Public Web Archives
- Session 6A: Rich and Krestel, My Approach - Your Apparatus? Entropy-Based Topic Modeling Domain-Specific Text Collections

The tweet also shows engagement metrics: 11 replies, 12 retweets, and 12 likes.

Why do journalists use web archives?

View unavailable page

- Deletion - page
- Site unavailable on live web

View change over time

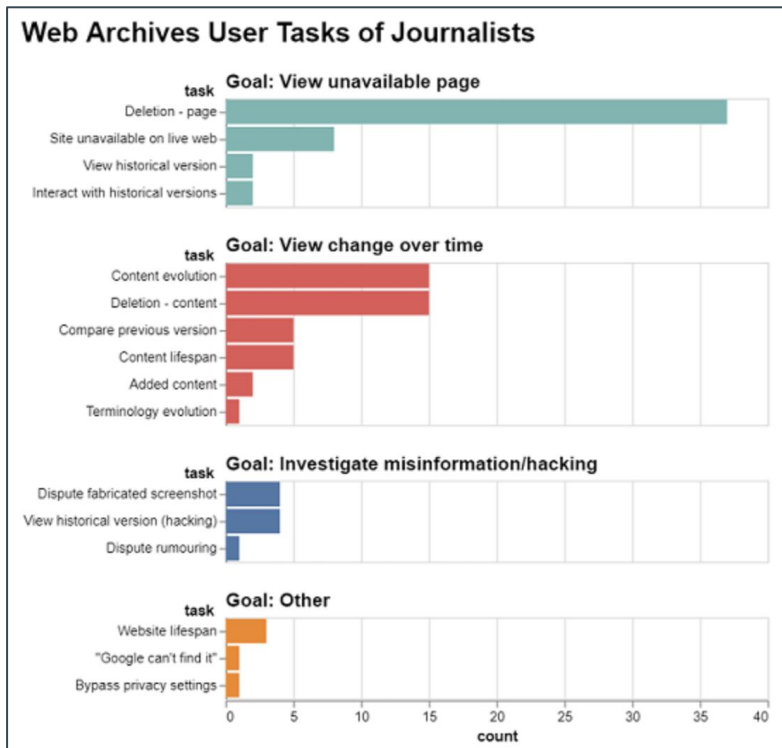
- Content evolution
- Deletion - content

Investigate misinformation/hacking

- Dispute fabricated screenshot
- View historical version

Other

- Website lifespan



Lesley Frew, Web Archiving in Popular Media II: User Tasks of Journalists, <https://ws-dl.blogspot.com/2022/08/2022-08-04-web-archiving-in-popular.html>

Searching Web Archives for Changed Text

added term/phrase	▼	"non cloth"
domain	▼	any
		Search
Search results for added phrase: non cloth		
NumFound: 2		
title	Novel Coronavirus (COVID-19) Department of Health	
url	https://coronavirus.health.ny.gov/node/1	
addition	2022-03-09 13:12:20	
<hr/>		
title	Novel Coronavirus (COVID-19) Department of Health	
url	https://coronavirus.health.ny.gov/home	
addition	2022-02-11 21:20:50	

- Only a few web archives offer full-text search.
- None (yet) allow you to search for what's changed.

- All travelers, and , should continue to follow all CDC travel requirements requirements.
- New Yorkers who don't feel well should stay home.
- Wear masks in public indoor space and wash hands regularly.

Last memento before the addition: <https://wayback.archive-it.org/4887/20220203045021/https://coronavirus.health.ny.gov/home>

First memento after the addition: <https://wayback.archive-it.org/4887/20220211212050/https://coronavirus.health.ny.gov/home>

Highlighting Politicians' Website Changes

deletion of the statement about George Santos's work history

George Anthony was then offered an exciting opportunity with Goldman Sachs but what he thought would be the pinnacle of his career was not as fulfilling as he had anticipated. George joined LinkBridge Investors, where he quickly advanced to become one of the youngest vice presidents in the industry.

George's work ethic comes from his mother, who came from nothing, but worked her way up to be the first female executive at a major financial institution. On September 11, 2001, George's mother was in her office in the South Tower. She survived the horrific events of that day, but unfortunately passed away a few years later.

Following in his mother's footsteps, George has a long history of philanthropic work across New York, and around the country. He has worked tirelessly with animal rescue efforts in the Tri-State area; and supports other organizations that help at-risk children and America's veterans.

George graduated from Baruch College with a bachelor's degree in economics and finance. He is fluent in Portuguese and Spanish. George lives on Long Island with his Husband, Matt; and their 4 dogs.

deletion of the statement about George Santos's grandparents being Jewish

George's grandparents fled Jewish persecution in Ukraine, settled in Belgium, and again fled persecution during WWII. They were able to settle in Brazil, where his mother was born. His father, who comes from Angolan roots, was also born in Brazil. Both his mother and father legally immigrated to the United States in search of the American dream, where they began their family on the foundations of life, liberty, and the pursuit of happiness. George's drive, commitment and determination were inspired by the legacy set and left by his mother. She gave up everything to provide for George and her family. She delivered the American dream to George, a debt that he wants to repay to the rest of his neighbors and constituents. George's mother was in her office in the South Tower on September 11, 2001, when the horrific events of that day unfolded. She survived the tragic events on September 11th, but she passed away a few years later when she lost her battle to cancer.

The American dream that George grew up living is now under an all-out attack from the Radical Left. Inflation is out of control and crime is rampant in the streets. New York City and Long Island used to be the gold standard for quality of life in America – now New Yorkers have watched their communities

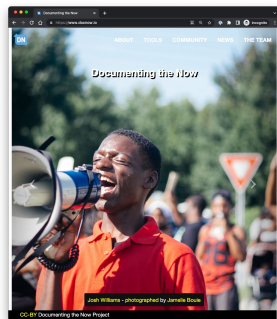
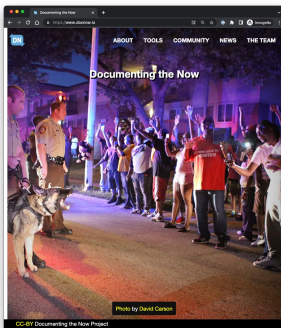
Lesley Frew, "Animating Changes in Webpages, Featuring George Santos's Biography", <https://ws-dl.blogspot.com/2023/02/2023-02-26-animating-changes-in.html>

Using Web Archives to Document Social Movements and Disinformation - Practice, Ethics, and Challenges

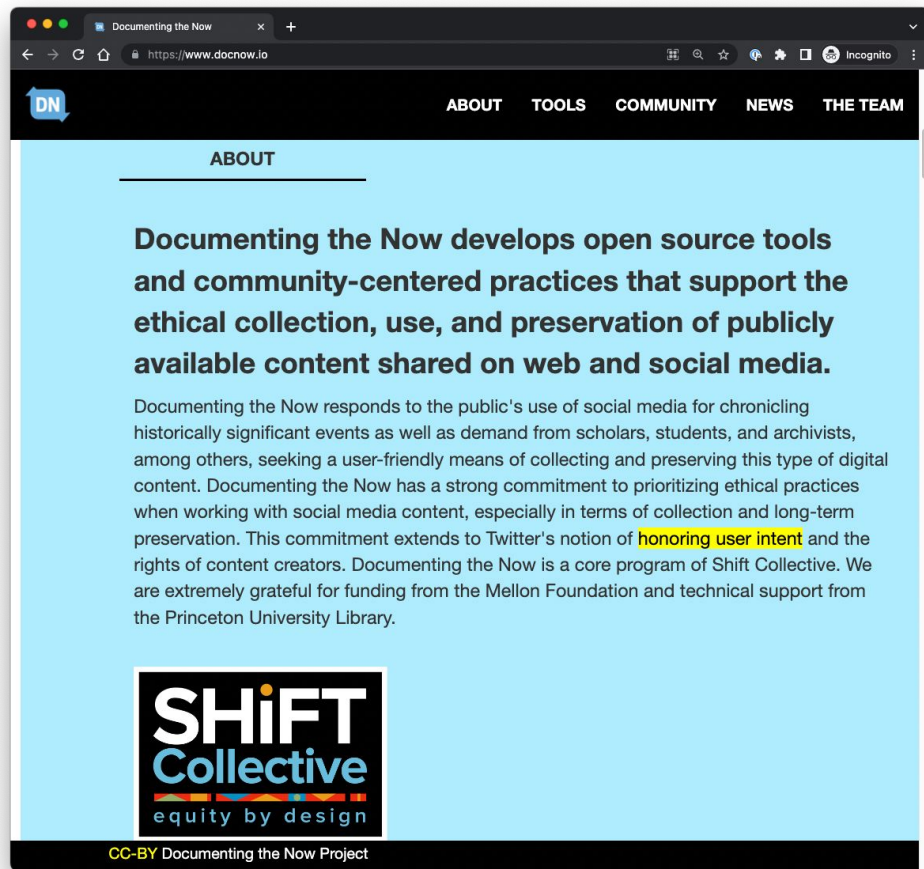
Using Web Archives to Document Social Movements and Disinformation - **Ethics**

Documenting the Now

- Emerged in 2016, in response to efforts to document social movements that were being recorded on social media
- Provides tools and best practices for archivists and researchers



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



National Forum on Ethics & Archiving the Web (EAW), 2018



Goal was to bring together
activists, librarians, journalists,
archivists, scholars, developers,
and designers

<https://eaw.rhizome.org/>

Michael Nelson, "Trip Report for the National Forum on Ethics and Archiving the Web (EAW)", 2018,
<https://ws-dl.blogspot.com/2018/04/2018-04-09-trip-report-for-national.html>

Who owns a user's posts on a social media platform?

- Users lack awareness of how social media platforms use their data *and* how that data can be used by third parties (including researchers and journalists)
- Survey found that a significant number of Twitter users were unaware their tweets could be used by third parties without consent (Fiesler and Proferes, 2018)

Bergis Jules and Ed Summers, "Documenting the Now Ethics White Paper", July 2018, <https://news.docnow.io/documenting-the-now-ethics-white-paper-43477929ea3e>, <https://www.docnow.io/docs/docnow-whitepaper-2018.pdf>

Michael Nelson, "Trip Report for the National Forum on Ethics and Archiving the Web (EAW)", 2018, <https://ws-dl.blogspot.com/2018/04/2018-04-09-trip-report-for-national.html>

How to handle potential disinformation or hate speech?

- Malicious actors can pretend to be regular users and spread disinformation
 - @Blacktivist Twitter account, created by Russian IRA, tweeted prolifically in support of BLM, captured as part of #blacklivesmatter research
- Amplification of disinformation and hate speech is a concern
 - Duke Univ opted to *not* digitize a collection from the Southern Poverty Law Center to avoid recirculating hate materials

Bergis Jules and Ed Summers, "Documenting the Now Ethics White Paper", July 2018, <https://news.docnow.io/documenting-the-now-ethics-white-paper-43477929ea3e>, <https://www.docnow.io/docs/docnow-whitepaper-2018.pdf>

Michael Nelson, "Trip Report for the National Forum on Ethics and Archiving the Web (EAW)", 2018, <https://ws-dl.blogspot.com/2018/04/2018-04-09-trip-report-for-national.html>

How to be inclusive without causing harm?

- Increased potential for harm for members of marginalized communities, especially when participating in activities such as protests or other acts of civil disobedience
- Activists in marginalized groups can face disproportionate harm from surveillance
 - researcher collecting tweets related to Ferguson was contacted for data by a company that provides "situational awareness" tools to law enforcement

Bergis Jules and Ed Summers, "Documenting the Now Ethics White Paper", July 2018, <https://news.docnow.io/documenting-the-now-ethics-white-paper-43477929ea3e>, <https://www.docnow.io/docs/docnow-whitepaper-2018.pdf>

Michael Nelson, "Trip Report for the National Forum on Ethics and Archiving the Web (EAW)", 2018, <https://ws-dl.blogspot.com/2018/04/2018-04-09-trip-report-for-national.html>

Ethical Conundrums and Budgeting

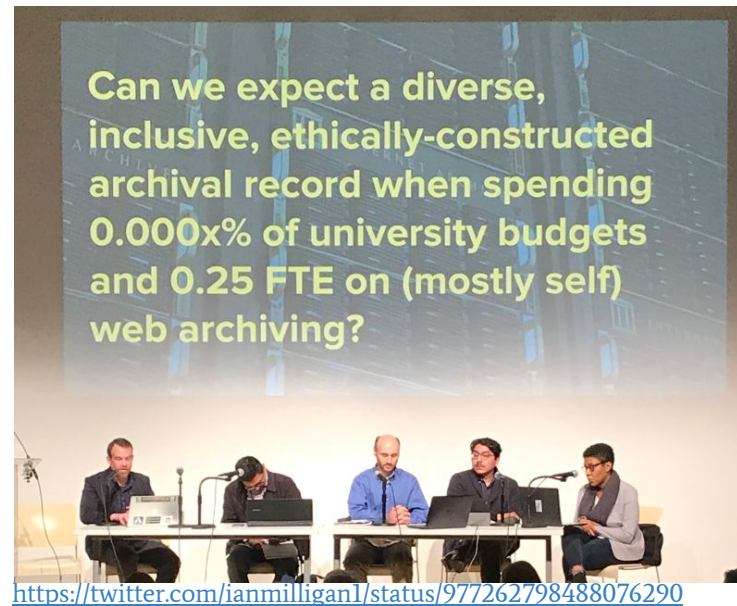


<https://eaw.rhizome.org/>

Ethical conundrums

- Ethical murkiness: What social media to proactively collect or not collect?
 - #MeToo: Important social movement, but women are sharing personal and traumatic experiences.
 - #Parkland / #NeverAgain: Important event and social movement, but involves minors.
- Each of these areas are important, but involve vulnerable individuals / communities or sensitive topics.
- Every decision not to collect results in a gap in historical record.

<https://twitter.com/ianmilligan1/status/977267978772320257>



<https://twitter.com/ianmilligan1/status/977262798488076290>

Michael Nelson, "Trip Report for the National Forum on Ethics and Archiving the Web (EAW)", 2018, <https://ws-dl.blogspot.com/2018/04/2018-04-09-trip-report-for-national.html>

Recommendations for Preserving Web/Social Media Content in Ethical Ways

- Engage and work with the communities being documented.
- Social media content should not be the only documentation collected.
 - *"Simply telling the story of what happens in social media is not enough, but it can be a beginning."*
- Follow social media platforms' TOS only where it lines up with the values of the communities being documented.
 - *"What is legal is not always ethical, and what is ethical is not always legal."*
- Apply traditional archival practices (appraisal, collection development, donor relations) to social media and web materials when possible.

Bergis Jules and Ed Summers, "Documenting the Now Ethics White Paper", July 2018, <https://news.docnow.io/documenting-the-now-ethics-white-paper-43477929ea3e>, <https://www.docnow.io/docs/docnow-whitepaper-2018.pdf>

Representation in Web Archives

- "For digital collections, who gets represented is closely tied with who:
 - writes the software
 - builds the tools
 - produces the technical standards
 - provides the funding"
- "The work we do as archivists, as librarians, as digital preservationists, have real consequences for marginalized people because who is remembered and how they're remembered dictates who gets violence perpetrated against them."
- "Neutrality is a threat to the legacies of marginalized people and by extension their lives."

Bergis Jules, "Confronting Our Failure of Care Around the Legacies of Marginalized People in the Archives", 2016, <https://medium.com/on-archivy/confronting-our-failure-of-care-around-the-legacies-of-marginalized-people-in-the-archives-dc4180397280>

Archiving the Black Web

<https://archivingtheblackweb.org/>



- Goal is to establish a "more equitable and accessible web archiving practice" to document the Black experience online
- Received \$2.5M grant from the Mellon Foundation
- Develop training program in web archiving aimed at documenting Black life, history, and culture

Using Web Archives to Document Social Movements and Disinformation - Practice, Ethics, and Challenges

Using Web Archives to Document Social Movements and Disinformation - **Challenges:** **Twitter**

We spent the Fall of 2020 studying archived Twitter

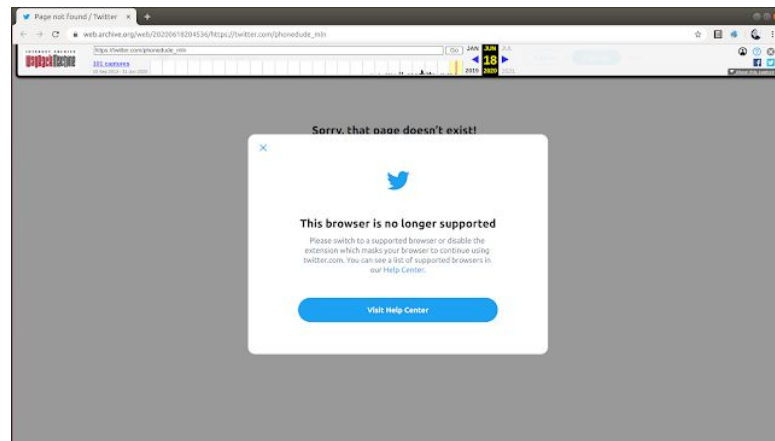
Kritika Garg, Himarsha R. Jayanetti, Sawood Alam, Michele C. Weigle, and Michael L. Nelson, "Replaying Archived Twitter: When your bird is broken, will it bring you down?," In ACM/IEEE JCDL 2021, pp. 160-169,
<https://arxiv.org/abs/2108.12092>

based on a series of blog posts:

Kritika Garg and Himarsha Jayanetti, "Twitter Was Already Difficult To Archive, Now It's Worse!", July 2020,
<https://ws-dl.blogspot.com/2020/07/2020-07-15-twitter-was-already.html>

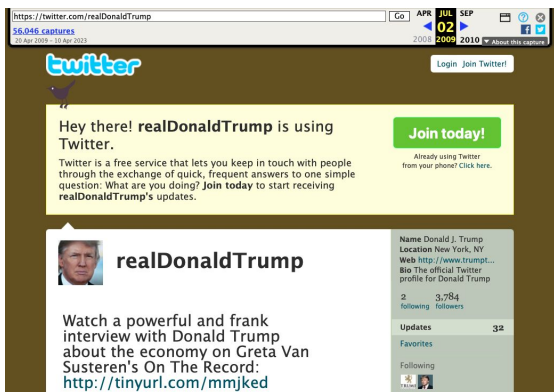
Himarsha Jayanetti and Kritika Garg, "New Twitter UI: Replaying Archived Twitter Pages That Never Existed", November 2020,
<https://ws-dl.blogspot.com/2020/11/2020-11-04-new-twitter-ui-replaying.html>

Kritika Garg and Himarsha Jayanetti, "Twitter Added Labels On Its Old User Interface", December 2020,
<https://ws-dl.blogspot.com/2020/12/2020-12-08-twitter-added-labels-on-its.html>



The live web keeps changing, so web archives must adjust to keep up

2009



<https://web.archive.org/web/20090702030955/https://twitter.com/realDonaldTrump>

2017

(Old UI)



<https://web.archive.org/web/20170702084625/https://twitter.com/realDonaldTrump>

2013



<https://web.archive.org/web/20130608234757/https://twitter.com/realDonaldTrump>

2020

(New UI)



<https://web.archive.org/web/202030407025620/https://twitter.com/realDonaldTrump>

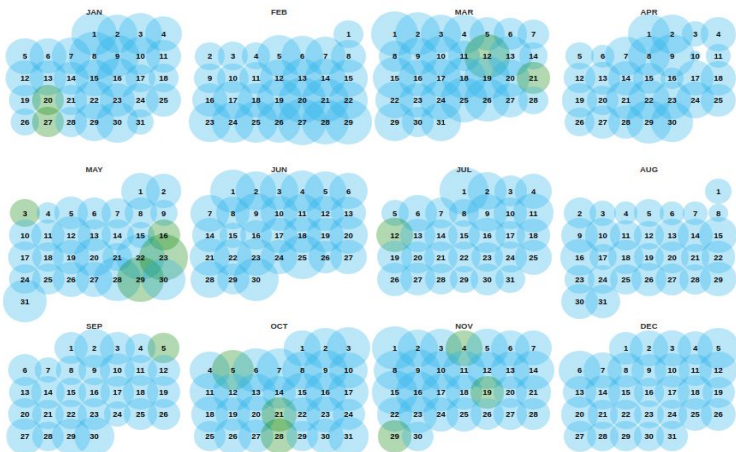
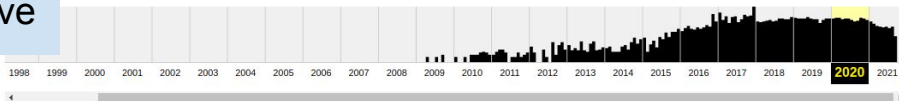
@realDonaldTrump Twitter account is well archived



Internet
Archive

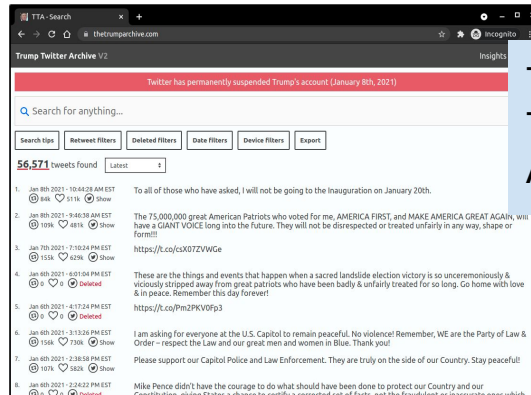
Calendar · Collections · Changes · Summary · Site Map

Saved 52,699 times between April 20, 2009 and September 8, 2021.



http://web.archive.org/web/20200701000000*/https://twitter.com/realDonaldTrump

Garg et al., ACM/IEEE JCDL 2021



The
Trump
Archive

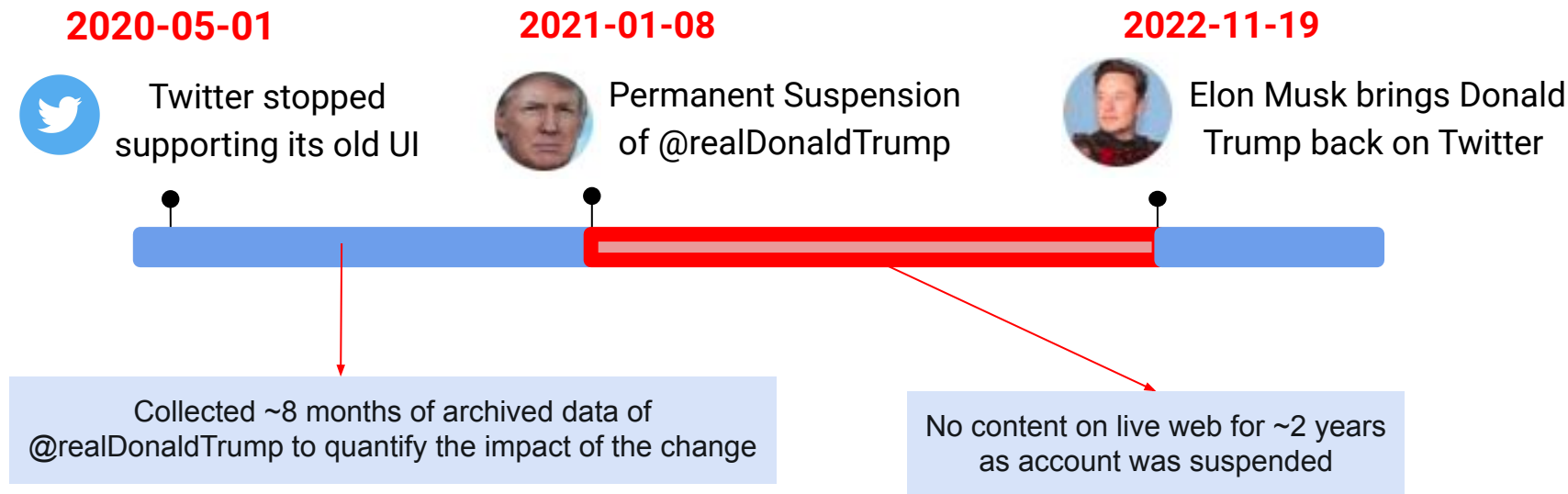
<https://www.thetrumparchive.com/>



Factba.se

<https://factba.se/trump/>

Used @realDonaldTrump to study the impact of Twitter UI change on web archives



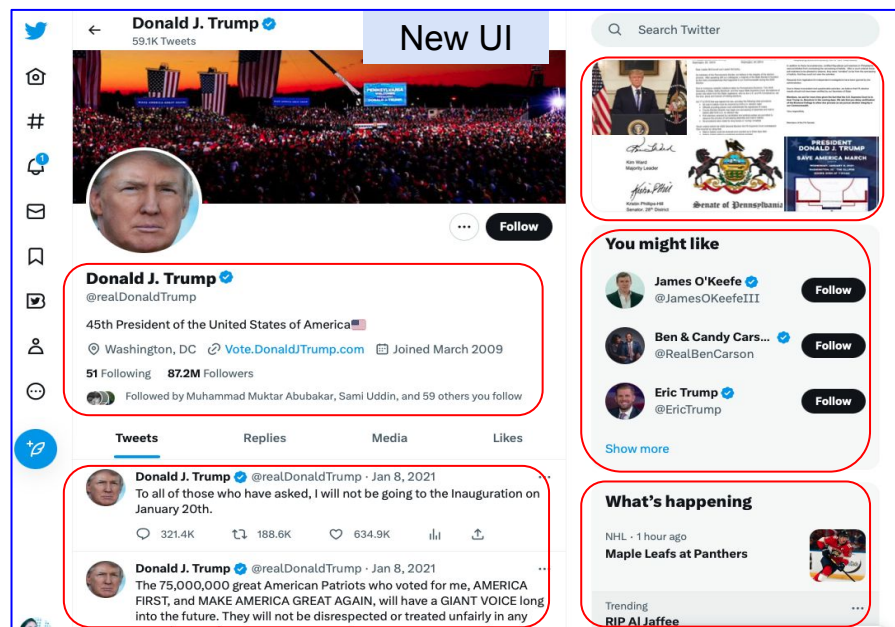
https://blog.twitter.com/en_us/topics/company/2020/suspension
https://en.wikipedia.org/wiki/Acquisition_of_Twitter_by_Elon_Musk

Tweets in old UI are embedded in HTML / New UI requires JSON requests



20 tweets and Twitter bio and are embedded in the base HTML

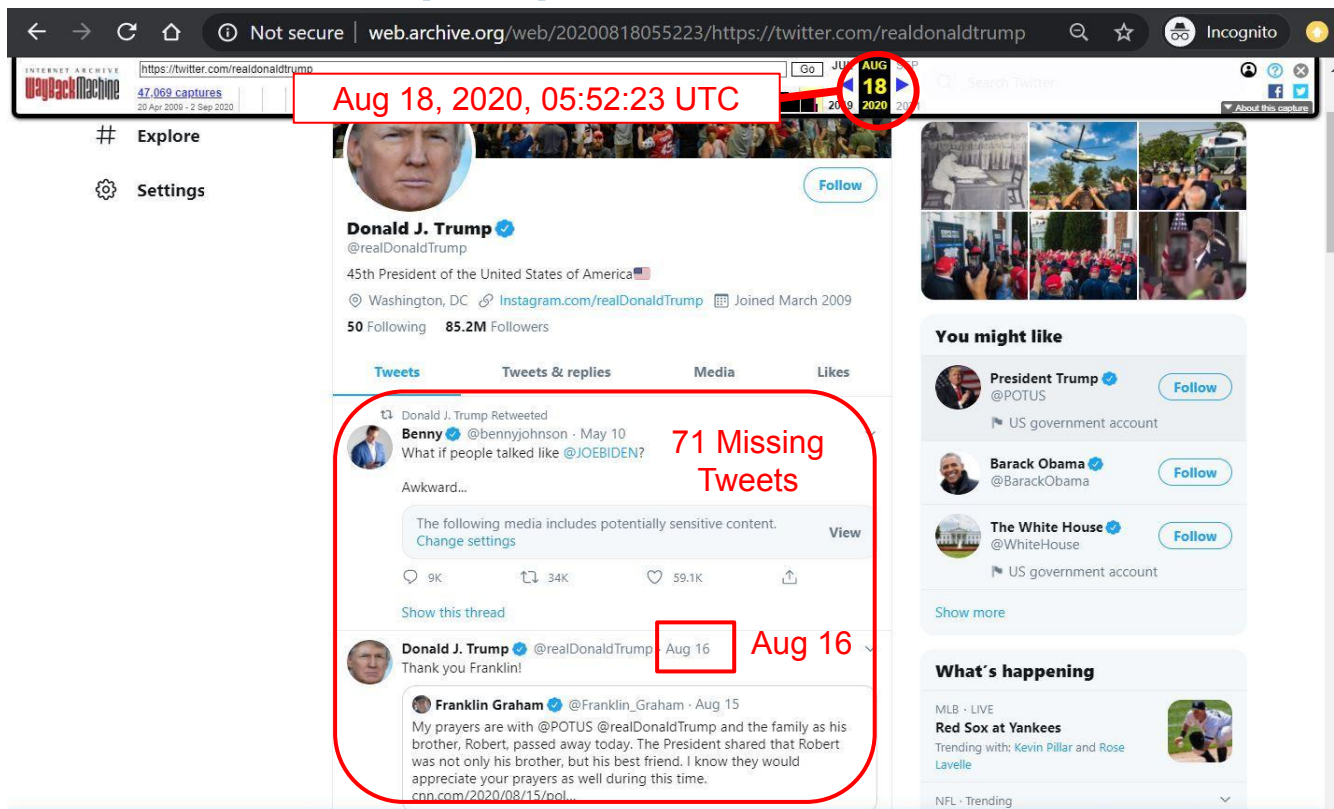
Content populated with follow-up XHR requests



base HTML contains only a skeleton, and all page sections are served dynamically through API JSON responses

(<https://api.twitter.com/2/timeline/profile/25073877.json?..>)

New UI mementos may replay pages that never existed on live web



<https://ws-dl.blogspot.com/2020/11/2020-11-04-new-twitter-ui-replaying.html>

Each section could have a different datetime

The image shows a screenshot of a web archive of Donald Trump's Twitter profile. The browser address bar shows the URL: <https://web.archive.org/web/20200818055223/https://twitter.com/realdonaldtrump>. The page is annotated with several colored boxes and labels indicating different datetimes for different sections:

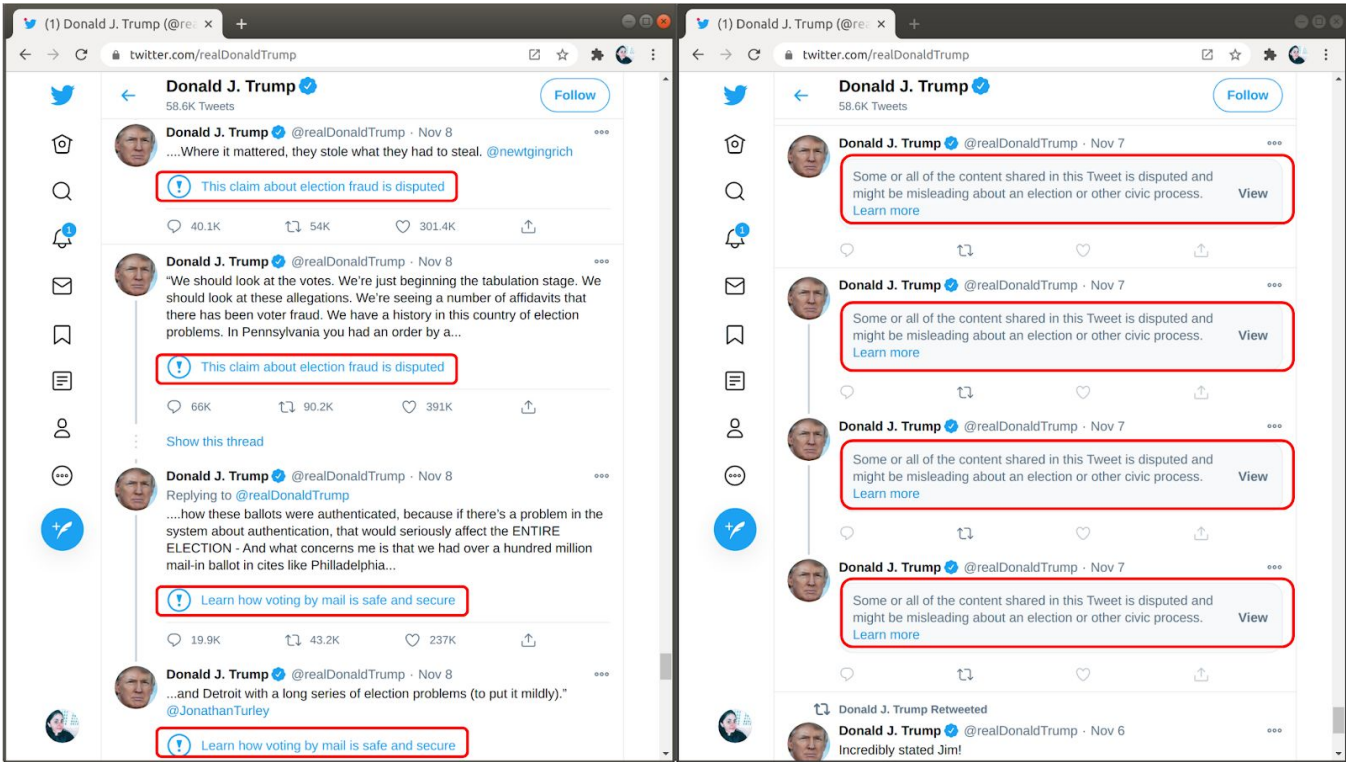
- Bio**: -1 day 5 hs 4 mins (blue box)
- Tweets**: -1 day 5 hrs 4 mins (red box)
- Media timeline**: -1 day 5 hrs 19 mins (orange box)
- You might like**: -1 day 5 hrs 19 mins (green box)
- What's happening**: -24 days 21 hrs 29 mins (magenta box)

Additional annotations include:

- A red box around the date "Aug 18, 2020, 05:52:23 UTC" in the browser's address bar.
- A red box around the date "Aug 18, 2020" in the browser's address bar.
- A red box around the text "71 Missing Tweets" in the tweets section.
- A red box around the date "Aug 16" in the tweets section.

Twitter introduced labels in May 2020

Were being
widely used by
November 2020



<https://ws-dl.blogspot.com/2020/12/2020-12-08-twitter-added-labels-on-its.html>

We categorized labels as “Fact-check” and “Violated Twitter Rules” (VTR)

Fact-check Labels



Donald J. Trump ✓
@realDonaldTrump

There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor of California is sending Ballots to millions of people, anyone.....

! Get the facts about mail-in ballots

8:17 AM · May 26, 2020 · Twitter for iPhone

36.1K Retweets 11.7K Quote Tweets 125.3K Likes



<https://twitter.com/realDonaldTrump/status/1265255835124539392>

Violated Twitter Rules Labels



Donald J. Trump ✓
@realDonaldTrump

This Tweet violated the Twitter Rules about spreading misleading and potentially harmful information related to COVID-19. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible. [Learn more](#)

Flu season is coming up! Many people every year, sometimes over 100,000, and despite the Vaccine, die from the Flu. Are we going to close down our Country? No, we have learned to live with it, just like we are learning to live with Covid, in most populations far less lethal!!!

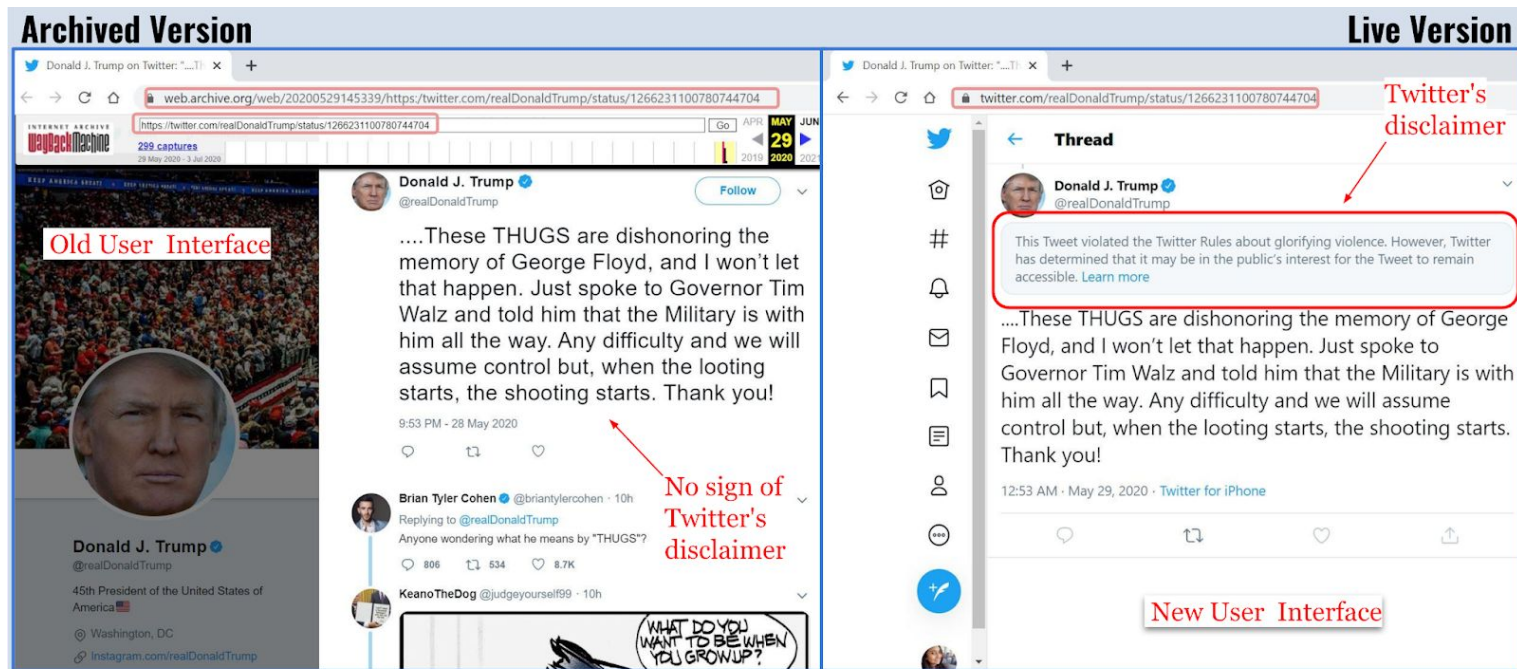
8:03 AM · Oct 6, 2020 · Twitter for iPhone

[View Quote Tweets](#)



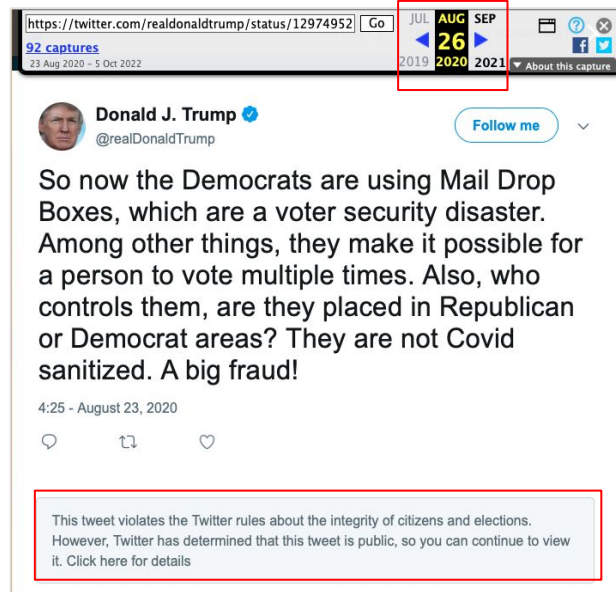
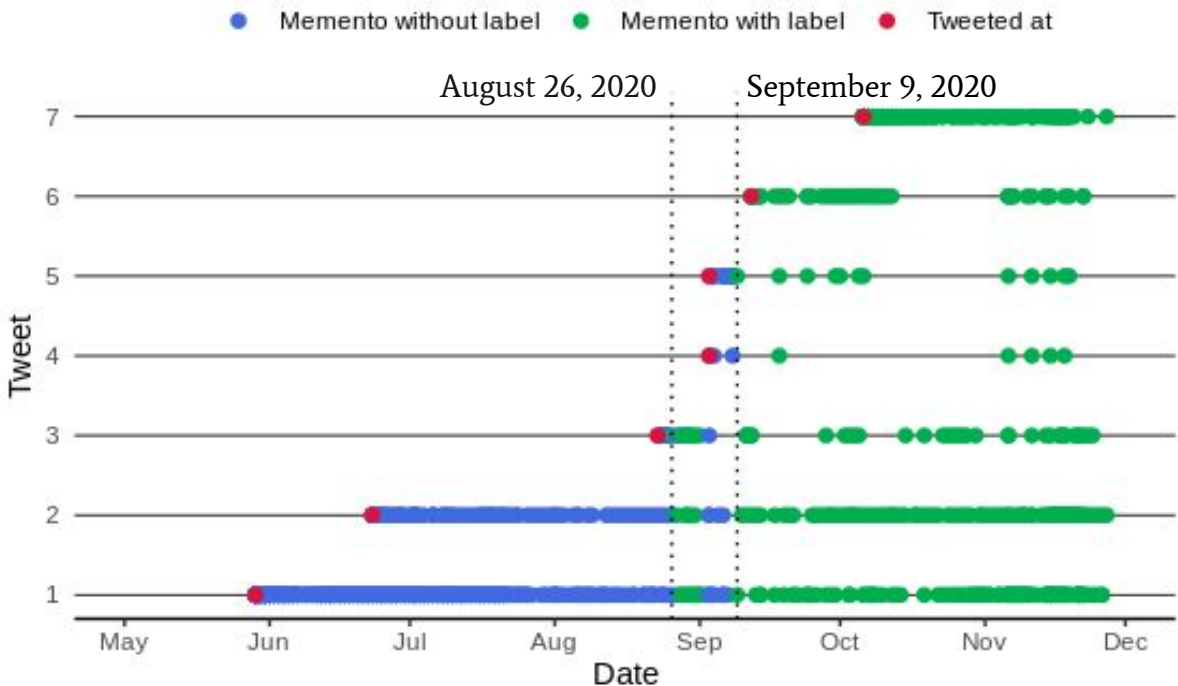
<https://twitter.com/realDonaldTrump/status/1313449844413992961>

Most web archives continued to archive the old UI, but labels were only included in Twitter's new UI



<https://web.archive.org/web/20200529145339/https://twitter.com/realDonaldTrump/status/1266231100780744704>

Twitter added VTR label to old UI by at least August 26, 2020



<https://ws-dl.blogspot.com/2020/12/2020-12-08-twitter-added-labels-on-its.html>

But, the “Fact-check” label is still not a part of the old Twitter UI



<https://web.archive.org/web/20210107122017/https://twitter.com/realDonaldTrump/status/1265255835124539392>



<https://web.archive.org/web/20211210011720/https://twitter.com/realDonaldTrump/status/1265255835124539392>

Today, the “Fact-check” label is no longer on the live web either



No Twitter's
Fact-check label

<https://web.archive.org/web/20221122044113/https://twitter.com/realDonaldTrump/status/1265255835124539392>

Using Web Archives to Document Social Movements and Disinformation - Practice, Ethics, and Challenges

Using Web Archives to Document Social Movements and Disinformation - **Challenges:** **Instagram**

Instagram was a significant front in the IRA's influence operation

- Russian Internet Research Agency (IRA)
- Manipulated social media narratives to influence American culture and politics in run-up to 2016 Presidential election
- Employed 1000+ to engage in influence operations, budget of over \$25M
- Analysis commissioned by the US Senate Select Committee on Intelligence

	Facebook	Instagram
Likes	37,627,085	183,246,348
Likes/Post	612	1,568
Comments	3,339,752	4,017,731
Comments/Post	54	34
Reactions	5,188,182	N/A
Shares	30,350,130	N/A
Total Engagements	76,505,149	187,264,079

New Knowledge, "The Tactics & Tropes of the Internet Research Agency", 2018, https://cdn2.hubspot.net/hubfs/4326998/ira-report-rebrand_FinalJ14.pdf

However, Instagram is not as studied as Twitter or Facebook

Twitter	Instagram
450 million monthly active users	2 billion monthly active users
8 million hits in Google Scholar	3 million hits in Google Scholar
Open (for now) API allows for searching and downloading public tweet content	API is focused more on allowing creation of content and monitoring engagement
Sharing based on retweets	No native sharing feature (though there are methods for "regrams")
Public posts and accounts viewable to not logged in users and bots	Public posts and accounts limited for not logged in users and bots

Instagram has banned accounts tied to misinformation

Facebook takes down Ukraine disinformation network and bans Russian-backed media

Meta says network ran websites posing as independent news entities and created fake personas

Facebook and **Instagram** have taken down a disinformation network targeting people in Ukraine, as their owner announced it was blocking access to the Russian state media outlets RT and Sputnik across the European Union.

<https://www.theguardian.com/technology/2022/feb/28/facebook-takes-down-disinformation-network-targeting-ukraine-meta-instagram>

Nearly 1,000 Facebook, Instagram accounts deactivated over COVID-19 misinformation

By Jordan Smith | Published December 2, 2021 | Updated 11:07AM | Social media | FOX TV Digital Team | ➔

Meta deleted nearly 1,000 Instagram and Facebook accounts, groups in several countries last month, most of which were found to belong to a sprawling network from China trying to spread COVID-19 misinformation.

The deactivated accounts, originating in Palestine, Poland, Belarus and China, violated Meta's Coordinated Inauthentic Behavior (CIB) policy, which is when users manipulate public debate for a strategic goal where fake accounts are central to the operation.

<https://www.fox29.com/news/nearly-1000-facebook-instagram-accounts-deactivated-over-covid-19-misinformation>

Instagram meme account with 14 million followers banned for coronavirus scam

No, PayPal is not sending people \$1,000.

By [Matt Binder](#) on March 30, 2020



<https://mashable.com/article/instagram-bestmemes-coronavirus-scam>

RFK Jr. kicked off Instagram for vaccine misinformation

By The Associated Press February 11, 2021

Instagram on Wednesday banned Robert F. Kennedy Jr., son of former presidential candidate Robert F. Kennedy, from repeatedly posting misinformation about vaccine safety and COVID-19.

Kennedy Jr. has amassed a huge following on social media, where he frequently posts debunked or unproven claims about vaccines. He also uses his social media pages to post about large pharmaceutical firms and environmental health concerns.

<https://apnews.com/article/rjf-jr-banned-instagram-vaccine-misinfo-d72913572f04ec9b23cec58881bc32d9>

For instance, Instagram has banned 10 of the "Disinformation Dozen"



1. Joseph Mercola
2. Robert F. Kennedy, Jr. **BANNED**
3. Ty and Charlene Bollinger **BANNED**
4. Sherri Tenpenny **BANNED**
5. Rizza Islam **BANNED**
6. Rashid Buttar **BANNED**
7. Erin Elizabeth **BANNED**
8. Sayer Ji **BANNED**
9. Kelly Brogan
10. Christiane Northrup **BANNED**
11. Ben Tapper **BANNED**
12. Kevin Jenkins **BANNED**

Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows

Updated May 14, 2021 · 11:48 AM ET ©

<https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes>

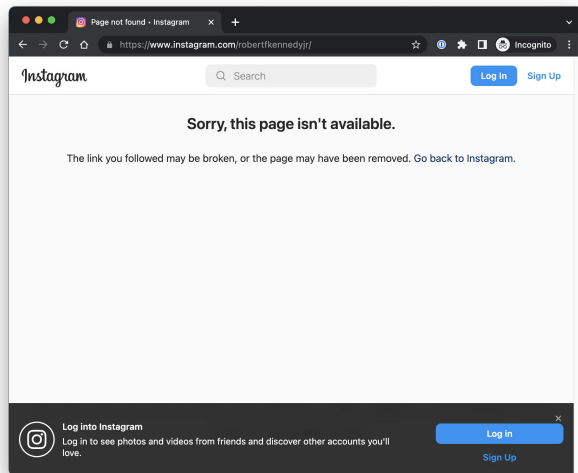
Majority of Covid misinformation came from 12 people, report finds

CCDH finds 'disinformation dozen' have combined following of 59 million people across multiple social media platforms

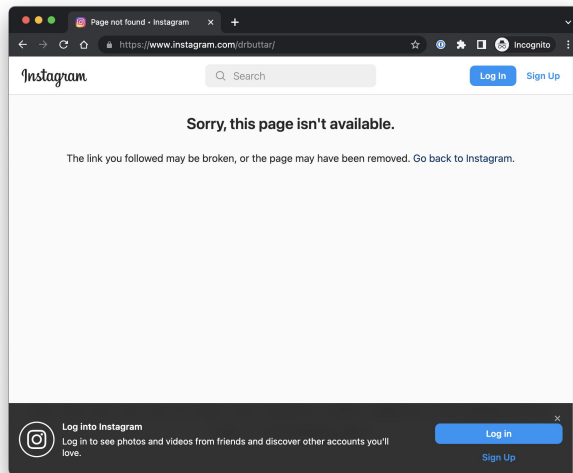
<https://www.theguardian.com/world/2021/jul/17/covid-misinformation-conspiracy-theories-ccdhd-report>

Center for Countering Digital Hate, "The Disinformation Dozen: Why Platforms Must Act on Twelve Leading Online Anti-Vaxxers", 2021, <https://counterhate.com/research/the-disinformation-dozen/>

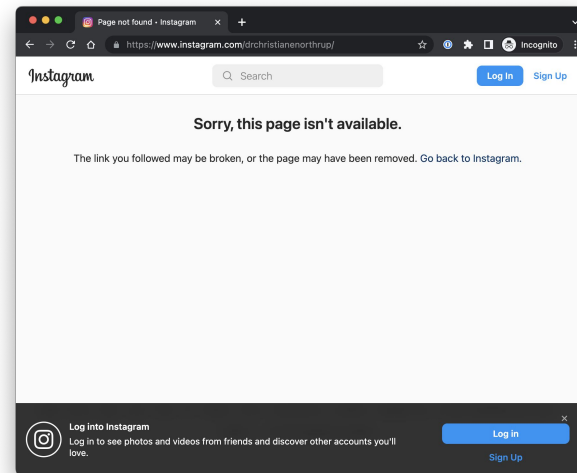
You can't view banned accounts on the live web



instagram.com/robertfkennedyjr/

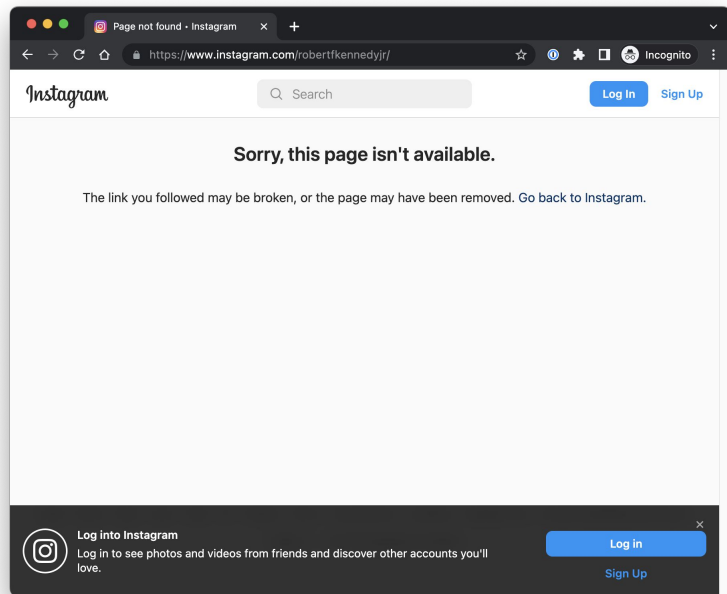


instagram.com/drbuttar/

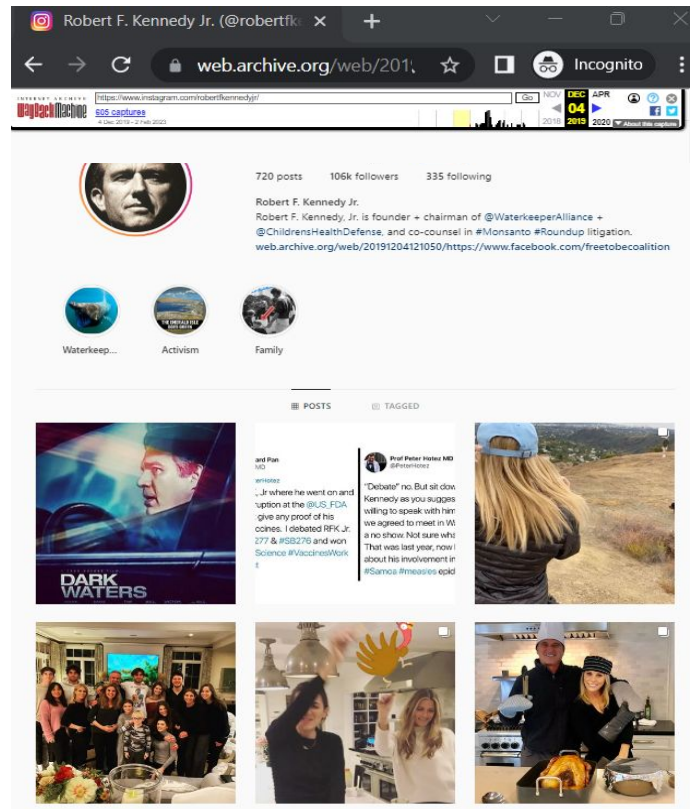


instagram.com/drchristianenorthrup/

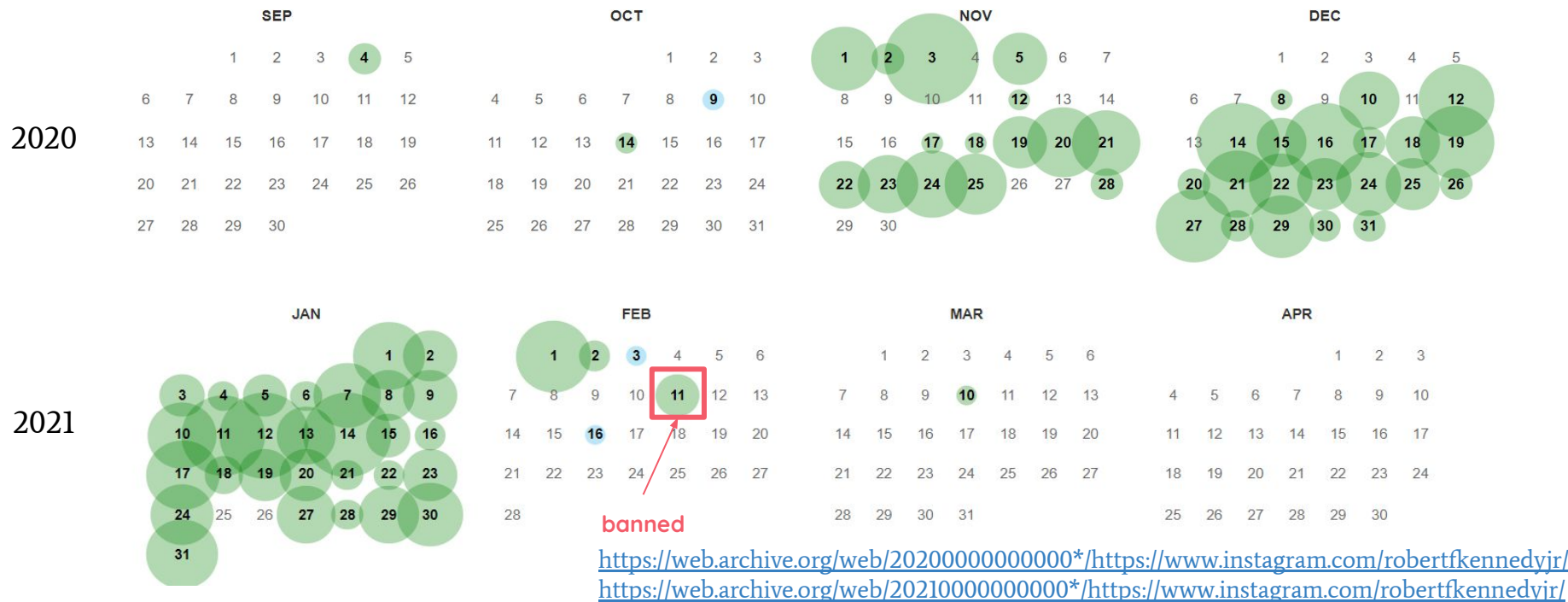
Need web archives to examine content before accounts banned



instagram.com/robertfkennedyjr/

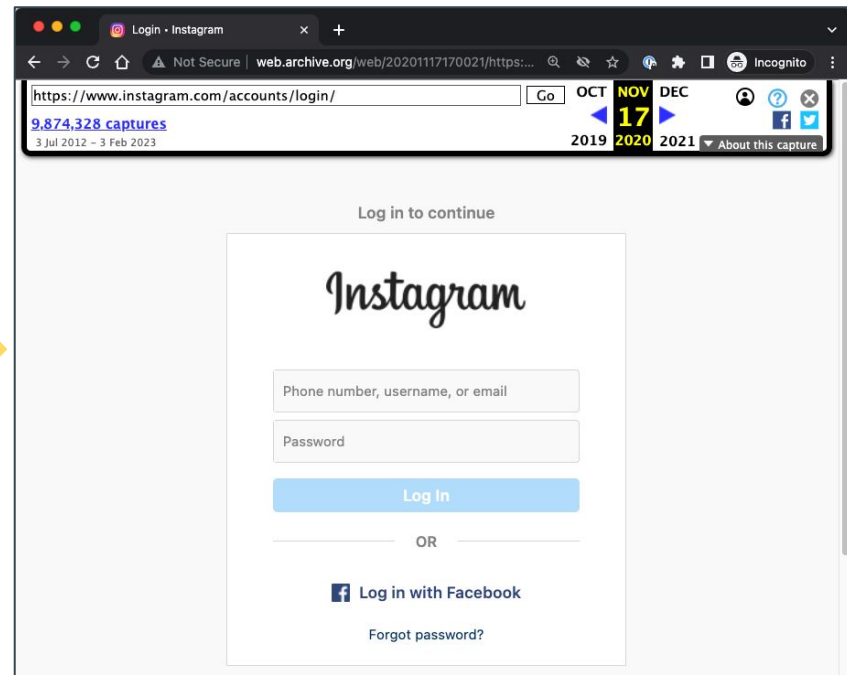
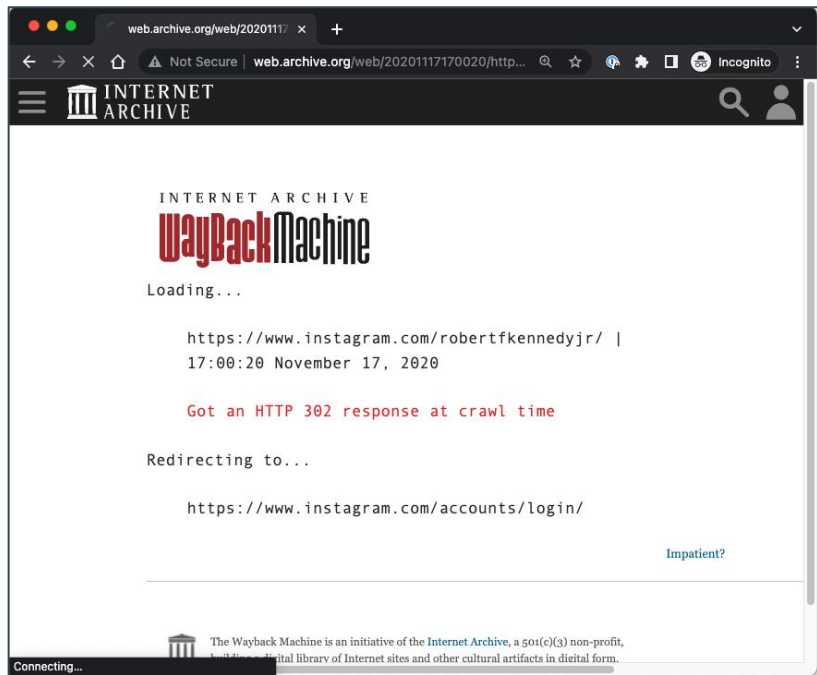


598 archived IG account pages for RFK, Jr. in IA



Haley Bragg, Himarsha Jayanetti, Michael L. Nelson, and Michele C. Weigle, Less than 4% of Archived Instagram Account Pages for the Disinformation Dozen are Replayable, ACM/IEEE JCDL 2023, based on REU report <https://arxiv.org/abs/2301.09188>

But most of these just redirect to the Instagram Login Page



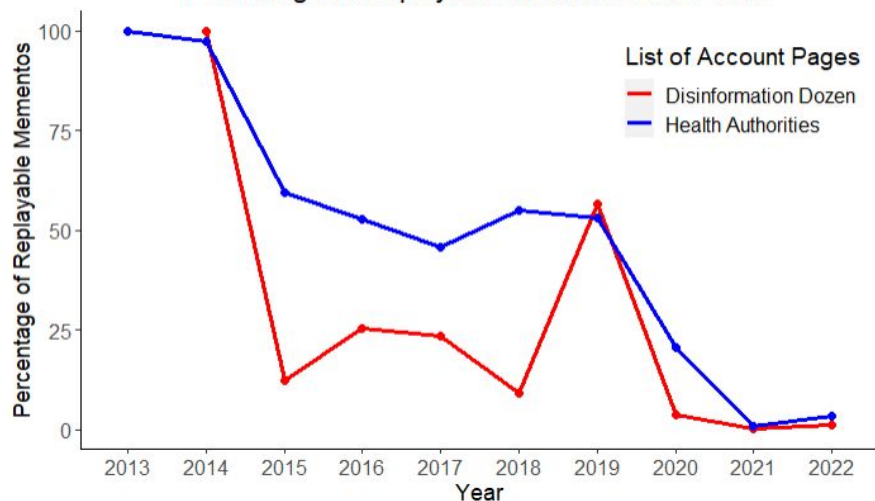
Only 6 successful replays for [instagram.com/RobertFKennedyJr](https://www.instagram.com/RobertFKennedyJr)

	Mementos	Percentage
Redirects to Login Page	592	99.0%
Replayed Account Pages	6	1.0%
Total Mementos	598	100%

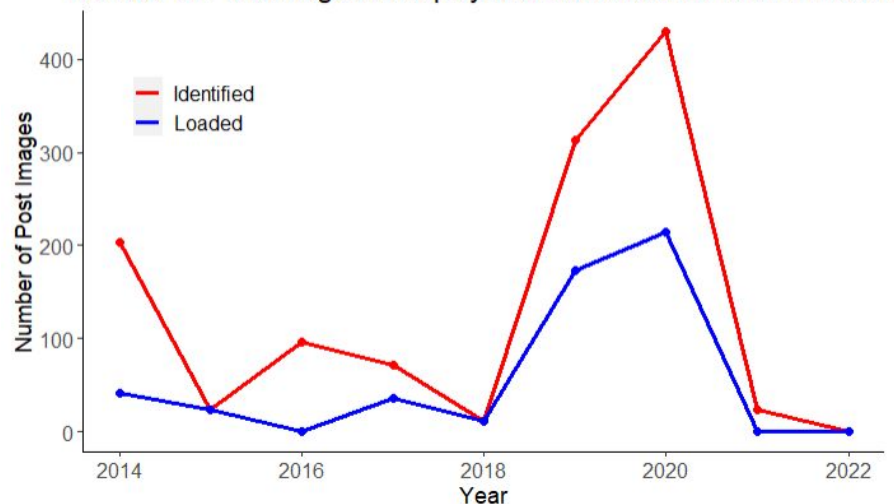
Instagram has made it even harder to archive pages

2021, 2022: Low percentages of replayable mementos and number post images available in those replayable mementos

Percentage of Replayable Mementos Over Time



Number of Post Images in Replayable Mementos for Disinfo Dozen



Bragg et al., ACM/IEEE JCDL 2023

Using Web Archives to Document Social Movements and Disinformation - Practice, Ethics, and Challenges

- Web archiving is an important tool for documenting events in our society.
- Web archives have been used to help document social movements, important cultural events, and disinformation on the web.
- Developers and archivists need to collaborate with activists and members of affected communities to ensure there is representation, but not harmful exposure.
- Changes in social media platforms and web technologies present multiple challenges for web archives.



<https://ws-dl.blogspot.com>

<https://weiglemc.github.io/talks/>

<https://bit.ly/weigle-lucy-2023>