

The Ferguson Principles

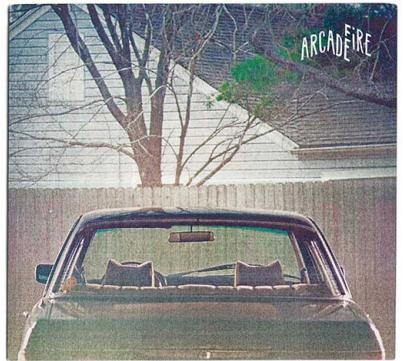
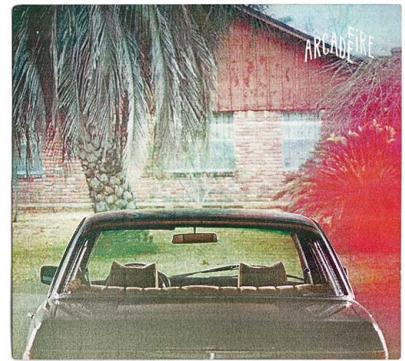
Social Media Archiving Praxis

Ed Summers
edsu@umd.edu

Bergis Jules
bergis.jules@shiftdesign.org











'Tsunami of data' could consume one fifth of global electricity by 2025

Billions of internet-connected devices could produce 3.5% of global emissions within 10 years and 14% by 2040, according to new research, reports [Climate Home News](#)



▲ A Google data centre. US researchers expect power consumption to triple in the next five years as one billion more people come online in developing countries. Photograph: Google/Rex

*We have a tsunami of data approaching.
Everything which can be digitalised is being
digitalised. It is a perfect storm.*

Anders Andrae



The Dr. Tressie McMillan Cottom ✅
@tressiemcphd

Following



Funny cause it's true



9:26 AM - 10 Aug 2018

140 Retweets 1,468 Likes



18

140

1.5K



How much of the web was in the Internet Archive...in 2016?

510 billion captures / 130 trillion URLs = 0.39%

1. Goel, V. (2016). [Defining Web pages, Web sites and Web captures](#). Internet Archive.
2. Schwartz, B. (2016). [Google's search knows about over 130 trillion pages](#). Search Engine Land.





August 10, 2014 — August 27, 2014
Ferguson: Michael Brown Killing
13,238,863

The shooting of Michael Brown occurred on August 9, 2014, in Ferguson, Missouri, a suburb of St. Louis. Brown, an 18-year-old black man, was fatally shot by Darren Wilson, 28, a white Ferguson police officer. This collection contains 13,238,863 tweets mentioning Ferguson from August 10, 2014 to

[Start Recording](#)

Source: [Wikipedia](#)

0:00 / 1:07

2014

AUG OCT DEC FEB APRIL JUNE

Ferguson: Michael Brown Killing 13,238,863

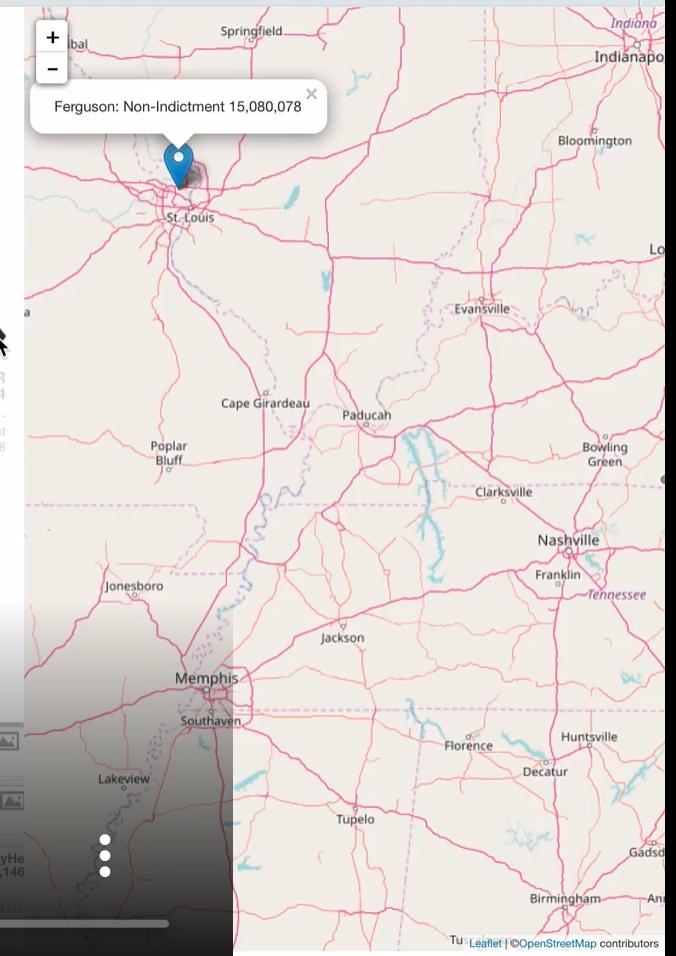
Ferguson: Non-Indictment 15,080,078

#FreddieG 2,983,934

Ferguson: Do Report 2,033, #Baltimore 2,195,3

#FreddieG 2,983,934

#FreddieG 2,983,934



TimeMapper: BlackLivesMatter Twitter Collection

```
% twarc search ferguson > tweets.jsonl
```

```
% twarc filter ferguson > tweets.jsonl
```

Tweet ID Datasets

Tweet ID Datasets

Twitter's [terms of service](#) don't allow tweet datasets to be published on the web, but they do allow tweet identifier datasets to be shared. This speaks to users rights as content creators, while also allowing researchers to share their data with others.

This site is a catalog of datasets that are publicly available on the web. If you would like to turn these tweet identifier datasets back into the original JSON first download the dataset and then use the [Hydrator](#) desktop application, or [Tward](#) if you are comfortable working at the command line.

You can add your own datasets to the catalog by following [these instructions](#). If you'd like updates when datasets are added please subscribe to the [RSS feed](#). All metadata listed here is licensed [CC0](#). You may want to refer to [our code of conduct](#) if you have questions or concerns about the datasets we list here.

iMCD Twitter Data

Creator: Urban Big Data Cen

Tweets: 65,000,000

Published: 2018-10-10

Date Coverage: 2014-07-01 - 2015-11-30

Tags: Glasgow, Scotland, Politics

► Description

Hurricane Florence Twitter Dataset

5430122 Creator: Mark Phillips

Tweets: 4,971,575

Published: 2018-10-08

0643077 Date Coverage: 2018-09-05 - 2018-10-05

Tags: hurricane, disaster, flooding, North Carolina, South Carolina

► Description

Aretha Franklin

Creator: Bergis Jules, Ed Summer

Tweets: 4,164,570

Published: 2018-09-05

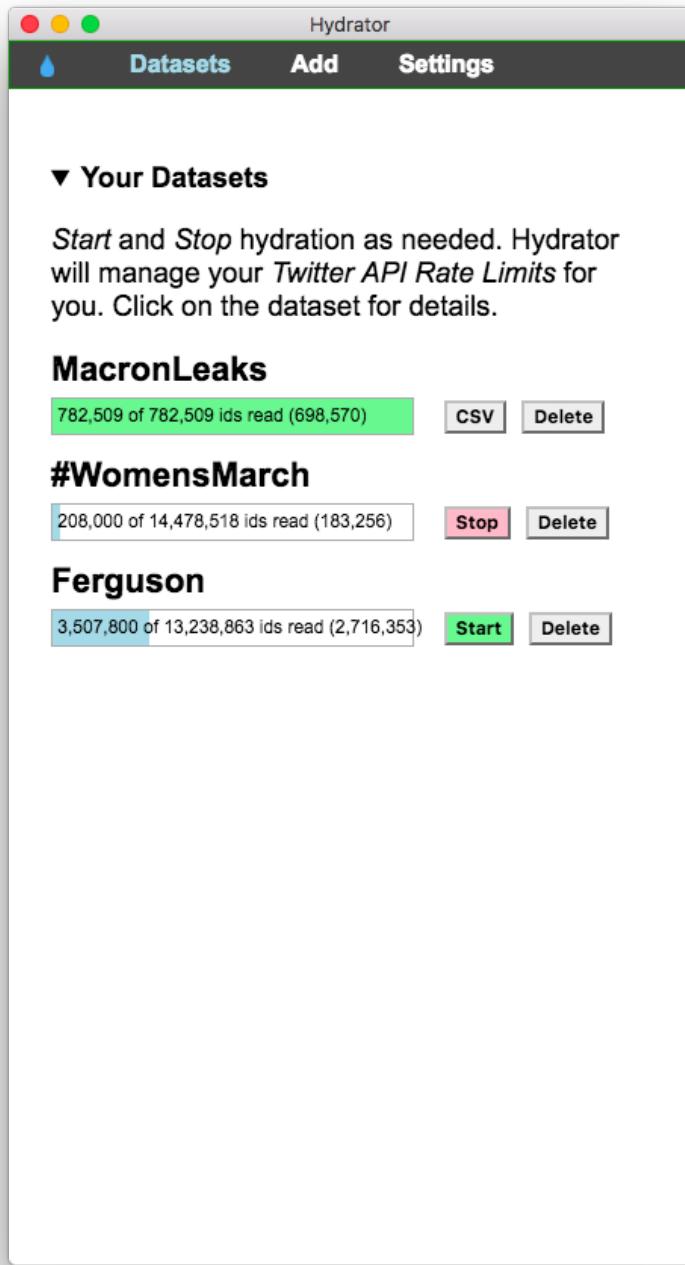
Date Coverage: 2018-08-08 - 2018-09-08

Tags: Music, Civil Rights Movement, African Americans

► Description

115th U.S. Congress Tweet Ids

Creator: Justin Littman, Yonah Bromberg Grabe

A screenshot of the Hydrator application window. The title bar says "Hydrator". The menu bar has "Datasets", "Add", and "Settings". Below that is a section titled "▼ Your Datasets". It contains three entries: "MacronLeaks" (status: Hydrating), "#WomensMarch" (status: Stopped), and "Ferguson" (status: Hydrating). Each entry shows the number of tweet IDs read and provides "CSV" and "Delete" buttons.

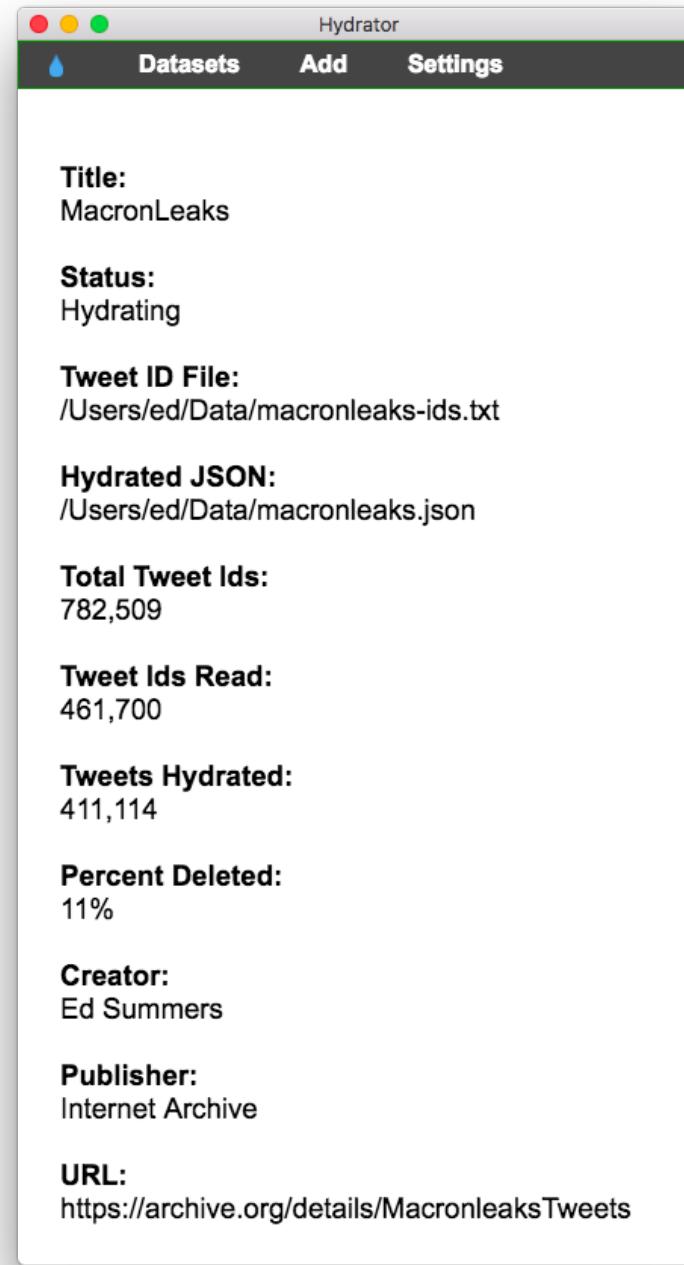
▼ Your Datasets

Start and Stop hydration as needed. Hydrator will manage your Twitter API Rate Limits for you. Click on the dataset for details.

MacronLeaks
782,509 of 782,509 ids read (698,570) **CSV** **Delete**

#WomensMarch
208,000 of 14,478,518 ids read (183,256) **Stop** **Delete**

Ferguson
3,507,800 of 13,238,863 ids read (2,716,353) **Start** **Delete**

A detailed view of the "MacronLeaks" dataset from the previous window. It shows various metadata fields: Title (MacronLeaks), Status (Hydrating), Tweet ID File (/Users/ed/Data/macronleaks-ids.txt), Hydrated JSON (/Users/ed/Data/macronleaks.json), Total Tweet Ids (782,509), Tweet Ids Read (461,700), Tweets Hydrated (411,114), Percent Deleted (11%), Creator (Ed Summers), Publisher (Internet Archive), and URL (<https://archive.org/details/MacronleaksTweets>).

Title:
MacronLeaks

Status:
Hydrating

Tweet ID File:
/Users/ed/Data/macronleaks-ids.txt

Hydrated JSON:
/Users/ed/Data/macronleaks.json

Total Tweet Ids:
782,509

Tweet Ids Read:
461,700

Tweets Hydrated:
411,114

Percent Deleted:
11%

Creator:
Ed Summers

Publisher:
Internet Archive

URL:
<https://archive.org/details/MacronleaksTweets>

```
http://localhost:3000/?url=https://bitly.com/4kb77v
http://localhost:3000/?url=https://bitly.com/4kb77v
{
  canonical: https://www.youtube.com/watch?v=oHg5SJYRHA0,
  status: 200,
  content_type: "text/html; charset=utf-8",
  title: "RickRoll'D - YouTube",
  description: "http://www.facebook.com/rickroll1548 As long as trolls are still trolling, the Rick will never stop rolling.",
  image: https://i.ytimg.com/vi/oHg5SJYRHA0/hqdefault.jpg,
  publisher: "YouTube",
  keywords: [
    "Cotter548",
    "Shawn",
    "Cotter",
    "lol",
    "gamefaqs",
    "CE",
    "reddit",
    "rettoocs",
    "no",
    "brb",
    "afk",
    "lawl",
    "pwnt",
    "Rickroll",
    "Rickroll'd",
    "Rick",
    "Roll",
    "Duckroll",
    "Duck",
    "rick",
    "roll",
    "astley",
    "..."
  ],
  short: https://bitly.com/4kb77v,
  long: https://www.youtube.com/watch?v=oHg5SJYRHA0
}
```



[whitehouse_diff](#)
@whitehouse_diff

Following



President Donald J. Trump Announces Intent to Nominate Individuals to Key Administrat... wayback.archive.org/web/2018110814 ... → wayback.archive.org/web/2018110818

...

John Barsa of Florida, to be an Assistant Administrator of the United States Agency for International Development (Bureau for Latin America and the Caribbean).

Mr. Barsa currently serves as Principal Deputy Assistant Secretary of the Department of Homeland Security's Office of Partnership and Engagement (DHS OPE). Previously, he served as the Acting Assistant Secretary for OPE. Mr. Barsa served with the United States Army Reserves for nearly a decade and was a member of the 11th Special Forces Group and the 450th Civil Affairs Battalion. Prior to his military service, In the U.S. Congress, he worked for Representative Lincoln Diaz-Balart, focusing on foreign policy and national security issues. During the Bush Administration, Mr. Barsa served in external engagement roles within the National Aeronautics and Space Administration and the Department of Homeland Security. He is a graduate of Florida International University and is a native Spanish speaker.

Kate Marie Byrnes of Florida, to be Ambassador Extraordinary and Plenipotentiary of the United States of America to the Republic of Macedonia.

1:45 PM - 8 Nov 2018



[TRENDING](#) [EXPLORE](#) [SAVED SEARCHES](#)

#nodapl #dapl #nokxl



Loaded 1094 tweets from Apr 30 9:52 AM to May 1 7:05 AM from the Twitter Search API.

Mike Hudema @MikeHudema 8 hours ago
"Someone needs to explain to me why wanting clean drinking water makes you an activist, and why proposing to destroy water with chemical warfare doesn't make a corporation a terrorist" ~Winona LaDuke @JustinTrudeau cdnpoli bcpoli
#NoDAPL #NoKXL #StopKM #WaterIsLife
<https://t.co/E2YqBP5JQx>

243 352

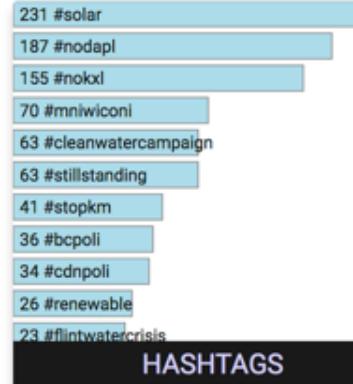
Mike Hudema @MikeHudema

TWEETS

Lily H @oceany 934 1,628 61,226
followers following tweets
@TheLeap_Org TheLeapManifesto #StopKM Environment, BernieSanders #NoDAPL @Mercola HealthRisingonFB #NoGMO RT's are not necessarily endorsements
Turtle Island 6 0.99 tw/hr

Barbara Navarro @BarbaraNavarro 10,518 9,777 80,366
followers following tweets

USERS



191 <https://buff.ly/2BsWhpK>
65 <https://www.npr.org/2018/04/29/606791799/all-these-farmers-are-fighting-the-dakota-access-pipeline>
26 <https://buff.ly/2q66Zds>
23 <https://www.indianz.com/News/2018/04/27/navajo-nation-leader-says-no-to-dakota-access-pipeline>
5 <http://www.kristv.com/story/38063997/judge-wouldnt-allow-dakota-access-pipeline>
5 <https://buff.ly/2HbmtaS>
4 <https://www.documentcloud.org/documents/444777.html>
3 <https://npr.org/2018/04/29/606791799/amid-key-new-rules-for-pipeline-safety>
3 <https://youtu.be/Onyk7guvHK8?a>
3 <https://buff.ly/2q7jVd>
3 <https://buff.ly/2r5pzL>
3 <https://www.youtube.com/watch?v=04MET-9ZHr>
2 <http://Www.christaorne.com>
2 <https://buff.ly/2HCNO2r>
2 https://mtlink.net/_s%253A%252F%252F.co%253B
2 https://mtlink.net/_s%253A%252F%252F.co%253B

URLS

"When we are asked how we use the water or how we view the Missouri River, we answer—Mni Wiconi. We have heard this term since we were little and it is the only way to describe how important the River is to us. We know how to live the land sustainably and the plants, trees, fish and animals which are given life by the water. We know how to take care of ourselves. But we cannot survive if the water is polluted."

Theo Iron Cloud Standing Rock Sioux Tribe

63 shares

"AS CARETAKERS OF OUR LAND, FAMILY FARMERS KNOW BEST WHAT'S GOOD FOR IT. WE STAND WITH THESE FAMILY FARMERS FIGHTING FOR THEIR LAND, LIVELIHOOD AND COMMUNITY."

-WILLIE NELSON

IMAGES

0:00 / 0:21

1 shares

VIDEO



Ferguson, Missouri (2017)

Ferguson Principles

1. Archivists must engage and work with the *communities* they wish to document on the web. Archives are often powerful institutions. Attention to the positionality of the archive vis-à-vis content creators, particularly in the case of protest, is a primary consideration that can guide efforts at preservation and access.

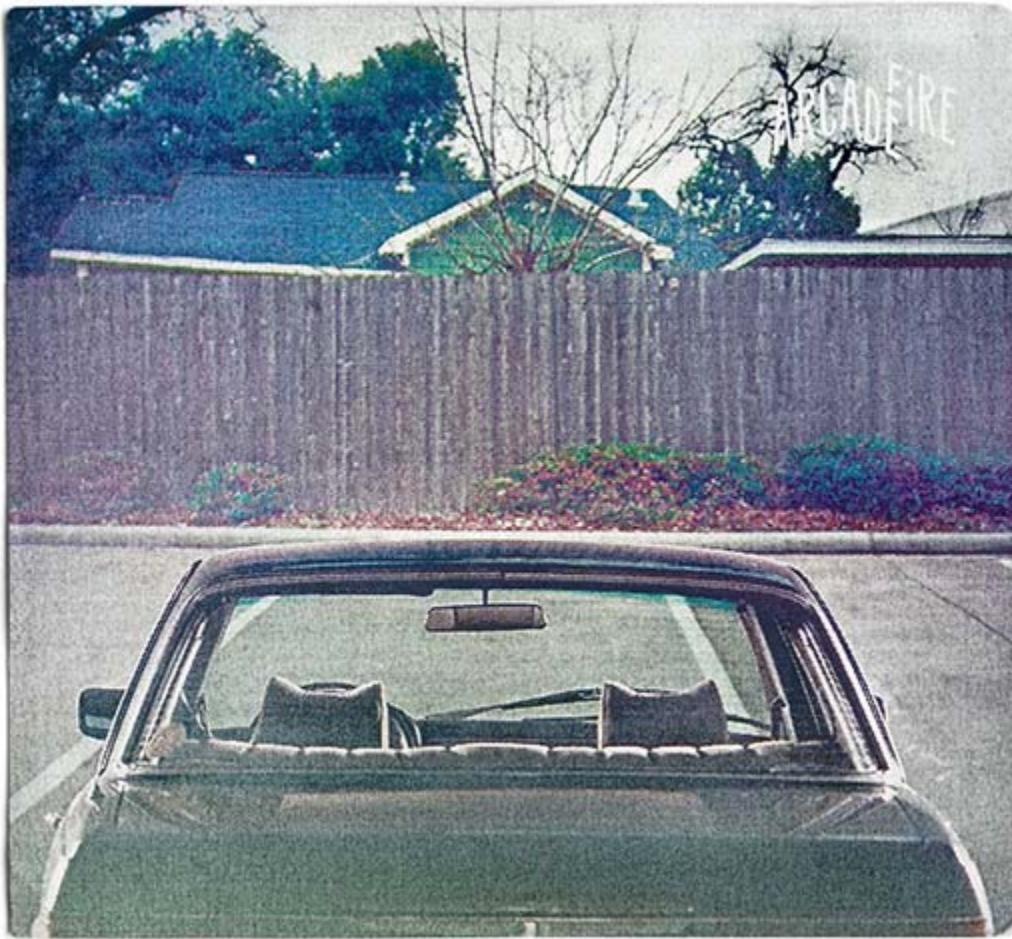
2. Documentation efforts must go beyond what can be collected without permission from the web and social media. Social media collected with the consent of content creators can form a part of richer documentation efforts that include the collection of oral histories, photographs, correspondence, and more. Simply telling the story of what happens in social media is not enough, but it can be a useful start.

3. Archivists should follow social media platforms' terms of service only where they are congruent with the values of the communities they are attempting to document. What is legal is not always ethical, and what is ethical is not always legal. Context, agency and (again) positionality matter.

4. Wherever possible, archivists should apply traditional archival practices such as appraisal, collection development, and donor relations to social media and web materials. It is hard work adapting these concepts to the collection of social media content, but they matter now, more than ever.



1. sustainable tech
2. activist community workshops
3. curriculum development



<https://www.docnow.io>