

Documenting the Now



ABOUT

Documenting the Now develops open source tools and community-centered practices that support the ethical collection, use, and preservation of publicly available content shared on web and social media.

Documenting the Now responds to the public's use of social media for chronicling historically significant events as well as demand from scholars, students, and archivists, among others, seeking a user-friendly means of collecting and preserving this type of digital content. Documenting the Now has a strong commitment to prioritizing ethical practices when working with social media content, especially in terms of collection and long-term preservation. This commitment extends to Twitter's notion of [honoring user intent](#) and the rights of content creators. Documenting the Now is a core program of Shift Collective. We are extremely grateful for funding from the Mellon Foundation and technical support from the Princeton University Library.



TOOLS

We build open source tools to help archivists, activists and researchers work with social media data. Follow us on [GitHub](#).

DocNow is a tool for appraising, collecting, and gathering consent for Twitter content. Features include:

- ✓ View trending topics across the globe in real time
- ✓ Live test and refine collecting parameters on recent tweets
- ✓ Explore content by users, media, URLs, and related hashtags
- ✓ Collect back in time via Search API and forwards with Stream API
- ✓ Share information about your collection with the public
- ✓ Gather consent from content creators
- ✓ Download Tweet ID archive for sharing following Twitter's terms of service

TRY IT!

twarc is a command line tool and Python library for collecting tweet data from Twitter's official API. It is designed for reliably collecting historical as well as realtime data, and can be used as a software library in your own tools and applications.

- ✓ Supports Twitter's v1 and v2 API endpoints
- ✓ Handles quota and API rate limiting
- ✓ Saves data using Twitter's official JSON representation for tweets
- ✓ Allows academic users to search back to the founding of Twitter in 2006
- ✓ Can be used for long running queries to collect millions of tweets
- ✓ Includes plugins for converting tweet data to CSV, network visualizations, and more

INSTALL

Hydrator is a desktop application for turning Tweet ID datasets back into tweet data to use in your research. It has been designed to be a reliable option for researchers who want to use their workstation for long running hydration jobs.

- ✓ Compliant with Twitter's user [intent policy](#)
- ✓ Cross-platform install (MacOS, Windows, Linux)
- ✓ Displays rehydration rate for collections
- ✓ Exports to JSON and CSV formats
- ✓ Resilient for long running hydration jobs
- ✓ Organizes datasets with extra metadata

DOWNLOAD

Social Humans is a label system for social media content. SH-C labels are for content creators to share their terms of consent for collection and re-use of their content, beyond Twitter's Terms of Service. SH-A labels are for academics and archivists who are collecting content to share contextual information about the collections. These labels are implemented in the DocNow application (and soon the Catalog), but can be used in any social media archiving program or project, including donor forms.

LEARN MORE

The Catalog is a community-sourced clearinghouse for tweet identifier datasets. Sharing tweet ids is a practice that is encouraged by Twitter as a way to share research data without negatively affecting users ability to have their data deleted or hidden from the web. We welcome your contributions!

- ✓ Describes how datasets were assembled, who collected them, and when
- ✓ Links to institutional repositories around the world where datasets are stored
- ✓ Allows users to easily upload their own dataset descriptions
- ✓ Datasets can be "hydrated" into Twitter data using twarc, Hydrator or other tools

VISIT

COMMUNITY

Building and maintaining a community of practice is a core tenet of the Documenting the Now project.

Conversation

Our **online community** has more than 600 active members, discussing topics around archiving social media: ethics, technical approaches, research, and practice in the field. [Join us](#) on Slack. Follow us at [@documentnow](#) on Twitter to follow our work and participate in hosted conversations by following [#docnowcommunity](#).

Workshops

Documenting the Now runs workshops with activist groups around the country, co-learning and providing training to address activists' needs and concerns in documenting their work and controlling their narrative. Past workshops have included partners Pettypropolis in Detroit, MI, Texas After Violence in Austin, TX, Don't Shoot PDX in Portland, OR, and the Southern California Library in Los Angeles, CA. If you are interested in conducting a workshop, contact us at info@docnow.io.

Support

Through our community outreach, we have learned that a primary need of activist communities is support in archiving their work. Archivists Supporting Activists (ASA) is a peer-matching network designed to help activists find archivists that hold themselves to a standard of non-extractive archival practice to support their archival needs. [Find an archivist](#) who can help you achieve your goals, or [sign up](#) to volunteer your skills.

Conferences

Documenting the Now has organized three meetings convening multidisciplinary researchers and practitioners to discuss the most pressing issues around social media and web archiving.

Most sessions were recorded and are available online:

- [2018 Ethics in Archiving the Web, New York, NY](#)
- [2017 Digital Blackness in the Archive, Ferguson, MO](#)
- [2016 Advisory Board Meeting, St. Louis, MO](#)

NEWS

The best way to follow Documenting the Now news is to follow our [Medium blog](#).

Press

06/2020 [Introducing the 2020-21 Data & Society Faculty Fellows](#): Documenting the Now Co-PI Meredith Clark to work with Data & Society.

06/2020 [Archivists Supporting Activists](#) launched by Documenting the Now in the wake of the murder of George Floyd to help coordinate memory work in activist groups who are demanding racial justice.

05/2020 [Historical Archives Once Silenced Marginalized Voices. Now Pandemic Archivists Want Them to Be Heard](#) by Marc Parry, The Chronicle of Higher Education.

05/2020 [“How Are We Going to Look Back on This Time?” Oral Historians Record Daily Life During COVID-19](#) by Molly Schwartz, Mother Jones.

04/2020 [Documenting COVID-19](#): a crowd sourced guide to Coronavirus documentation efforts initiated by Documenting the Now.

MORE PRESS

THE TEAM

Dr. Meredith Clark
Academic Lead, Co-Principal Investigator

Bergis Jules
Project Director, Co-Principal Investigator

Ed Summers
Technical Lead, Co-Principal Investigator

Zakiya Collier
Community Manager

Alexandra Dolan-Mescal
UX and Web Designer

Francis Kayiwa
DevOps Engineer

CONTACT US

info@docnow.io