

A Roundtable

THE ETHICS OF BORN-DIGITAL COLLECTING

*Featuring Alexandra Dolan-Mescal, Yvonne Ng,
Jessica Venlet, and Lauren Work*

March 22 | 10:00-11:30 AM EST



BitCurator
CONSORTIUM

Welcome!



- Introduction
- Speaker Presentations
- Discussion

<http://www.bitcuratorconsortium.org>

Use Case for Born Digital Collection at UVa Library

Lauren Work
@squaredsong
University of Virginia Library

I'm Lauren Work, and I'm the digital preservation librarian at the University of Virginia Library.

I'm going to talk a bit about an academic institutional use case for issues around ethics in born digital collecting.

I'll also frame the context of my own work to start as well: I'm a librarian and digital preservationist, but work very closely with archivists and am institutionally situated within the Special Collections, Archives & Preservation department of the library.

So my archivist colleagues and myself aim to center around of the existing ethics & values frameworks for our professions, but as we know, that these statements only go so far to guiding practice, so what I wanted to highlight over my presentation is how we are trying to build structural ethical ways we work so we can build and iterate on some of these ideas in practice.



So the specific use case around ethics in born digital collecting relates to the act of domestic terrorism that happened Charlottesville on August 11 & 12th of 2017.

I imagine many of you are well familiar with this act of terrorism where white supremacists organized and came to both the University of Virginia, where they were met by student activists and counter protestors, and then the next day to the city of Charlottesville.

Heather Heyer, a counter-protester to white supremacists was killed by a white supremacist who purposefully drove a car into a crowd and many more were injured.

During and after these events, a decision was made by the Dean of the Library that we should collect digital materials to try to center the UVa and Charlottesville community and local experience of the events.

- Twarc
- Webrecorder
- Google Forms
- Archive-It
- Omeka collecting website
- In-person outreach events

- Community work
- Student organization discussions
- Continued outreach
- Documentation and follow-through

On the left side of the slide are tools and systems we set up to collect - some of which we had in place prior to the events like Archive-It, webrecorder, and a Google form for website submissions, but other tools for data collection such as twarc and an omeka public collection website, were completely new. As many of you know, some of these tools make it easy to rapidly collect large amounts of data, which we'll come back to in the next slide

To the right side of the slide is a representation of the very necessary accompanying personal and community actions that we took around collection for the event. We held outreach events in the community to help people submit materials to the omeka site, we reached out and talked with student activist groups who opposed the white supremacists at UVa on Friday night, and worked to provide documentation and follow up with those interested in partnerships around collecting.

Thinking through frameworks for ethical decision making, policy, and actions over time

- Community work
- Student organization discussions
- Continued outreach
- Documentation and follow-through

And I wanted to talk a bit here about how we tried to structure privacy and ethics concerns within some of our work around collecting, including how it is integrated into tools for collection.

But more importantly, I also wanted to mention a few ways we are structuring and continuing, slower in-depth work around things like our institutional role (or not) in communities, collection policy, student organizations.

So a few things we did around collection included:

And to start, this was simply asking lots and lots of questions.

We talked to the folks involved in collection around the Documenting the Now, A People's Archive of Police Violence in Cleveland, Baltimore Uprising, and Documenting Ferguson

-Thinking very carefully around all aspects of what submitting to the digital collection site would mean for the community, for the library, and for the collection.

And this careful thinking lead to a slower launch of the website, which was a bit of a struggle for some in the library who wanted it released more quickly, but my own personal opinion is that it was worth the time to get it right and carefully think through

the issues.

"Unite the Right" Rally and Community Response

[Browse Items](#)[Search Items](#)[Contribute an Item](#)[About the Archive](#)

Contribution Terms of Service

You are being asked to contribute your recollections, photographic images, video, and/or other media to the University of Virginia (UVA), which is developing a permanent digital record of the events surrounding the Unite the Right rally and counter protests in Charlottesville, Virginia, in August 2017. You represent and warrant that the material you submit was created by you and was not copied from or based, in whole or in part, upon any other photographic, literary, or other material.

SUMMARY OF TERMS*

- You must be at least 18 years old.
- Submitted material must be owned and/or created by you.

Some what these structural ethical components looked like were things like ensuring that our information policy director has written and vetted a terms of service for submission that would clearly state for users:

- What it meant to submit digital materials to the collection site, and how these materials may be used or made public (or could be kept private)
- Allowed us to effectively transfer the terms of a record transfer/deed of gift to an online form that protected the rights of the submitter while also allowing the library to take collection management steps as necessary

We also focused on making materials available as quickly as possible (with permission) to the community through the Omeka site, so people could see their submissions and know it has not disappeared behind institutional walls.

We also talked with many community groups that were doing their own work of collection, and tried to see where their goals and mission and collection aligned with either skills or resources we could provide - not necessarily by us taking on any of their collection or ownership of their community work, but where we could partner with resources as desired.

What we're still working on:

I had mentioned that Twitter data collection was new for UVa, and for many administrative and community reasons, we have not made our data available yet.

On the administrative side, we had no existing policy, procedure, or infrastructure by which we could make tweets available.

On the ethical side, we needed to work through continuing to work with student groups around potentially getting their consent and donation of their own social media presence.

We also have to make decisions around what the cost benefit is for providing large sets of data around an event that continues to have ongoing litigation, and deep impacts on the local UVa and Charlottesville community, including threats of physical violence and doxxing, and the lingering trauma in the community. Again, I personally feel that this delay is ok, and that we should be asking ourselves and the community these questions.

The institutional reality in 2017 was also that we did not yet have a University Archivist, so her arrival in 2018 has also helped us begin to rebuild connections within UVa, particularly with student groups, as well as community centered archival work. These are important components to building out and sustaining relationships and trust.



Justin Littman

@justin_littman

Follow



Does embargoing the release of a Twitter dataset of a social movement / activism mitigate possible harm? Users have chance to delete tweets. Less likely to be of use to law enforcement / state security. Thinking as I ready our [#Charlottesville](#) tweet ids for release.

3:48 AM - 28 Feb 2018 from [Washington, DC](#)

3 Retweets 14 Likes



3



3



14



I wanted to take a minute here to also highlight people who are constantly asking these questions around ethics, practice, and communities, and how much this has helped with our work at UVa on this and other collections. Justin's question around embargos for the Charlottesville hash tag was a good one, as were the responses to how folks who work in this space daily are thinking through these issues.



Ed Summers @edsu · 28 Feb 2018

I don't think there is a hard and fast rule. But where possible it's important to connect with the communities that you are documenting and get a sense of what they want. cc/ @squaredsong



2



5



Texifter @texifter · 1 Mar 2018

It is not clear that a community could have a unitary preference so the notion of "checking" does not make sense. How would that work in a practical sense? When do movement participants all agree about anything this complex?



2



Justin Littman @justin_littman · 1 Mar 2018

Right. #Ferguson has a community. But how about #MeToo? #IranProtests? Definitely less clear / more challenging.



2



1



Deen Freelon @dfreelon · 1 Mar 2018

An embargo seems like a good precaution to me, especially if there's no pressing need to release immediately.



1



4



How might we build connections in communities that allow us to better understand and think about consent? How do you do this with huge global movements? Why do you feel the need (or does your institution feel the need) to immediately release information? Who might this harm if you do so? Who would it help, and why? What does time grant individuals and communities? Many of these questions are in an excellent social media collection development overview from GW in the resource doc: <https://gwu-libraries.github.io/sfm-ui/resources/guidelines>



Charlottesville Tweet Ids Version 1.0

Littman, Justin, 2018, "Charlottesville Tweet Ids", <https://doi.org/10.7910/DVN/DVLJTO>, Harvard Dataverse, V1

 Cite Dataset ▾

 Learn about [Data Citation Standards](#).

Description

This dataset contains the tweet ids of 7,665,497 tweets related to events in Charlottesville, Virginia in August, 2017. They were collected from the Twitter API using [Social Feed Manager](#).

These tweet ids are broken up into 2 collections. The collections are:

- Twitter search (charlottesville-search.txt): Search performed using the query "#charlottesville OR #standwithcharlottesville OR #defendCville OR #HeatherHeyer OR #UnityCville"
- Twitter filter (charlottesville-filter.txt): Filter stream using the filter "#charlottesville, #standwithcharlottesville, #defendCville, #HeatherHeyer, #UnityCville"

There is a README.txt file for each collection containing additional documentation on how it was collected.

The GET statuses/lookup method supports retrieving the complete tweet for a tweet id (known as hydrating). Tools such as [Twarc](#) or [Hydrator](#) can be used to hydrate tweets.

Per Twitter's Developer Policy, tweet ids may be publicly shared for academic purposes; tweets may not.

Questions about this dataset can be sent to sfm@gwu.edu. George Washington University researchers should contact us for access to the tweets.

[Feedback](#)

I wanted to end by highlighting where some social media materials are located currently around some of the Twitter data by GW, as they have had the policies and thinking around ethical data sharing in place. So while we work to get there ourselves at UVa, I wanted to highlight some great folks who have already done so.




Case Study: Born-digital Collecting in the Moment



Jessica Venlet

Asst. University Archivist for Digital Records
UNC-CH Libraries, Wilson Special Collections Library
March 22, 2019





UNC-CH Confederate Monument, known as Silent Sam
Photo credit: [Katie Sweeny, DTH, 2011](#)



The Silent Sam pedestal the day after the statue was toppled.
Photo credit: me, 8/22/2018



Pedestal removed, turf renovation in progress.
photo credit: me, 3/21/2019

Can & should we collect tweets?

#SilentSam

#SilenceSam

#StrikeDownSam

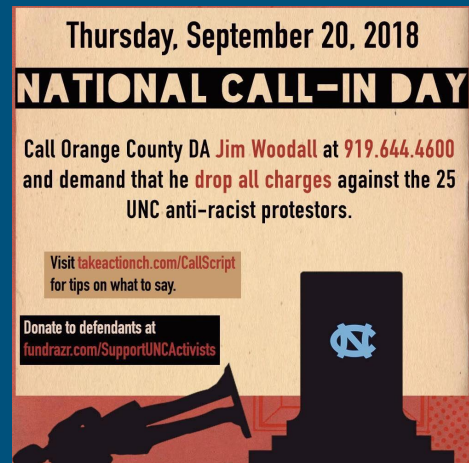
#BoycottUNC

In between all or nothing

- Snapshots rather than vacuum
- Time bound collecting
- Twarc
- Hashtag searches only
- Ask for permission for Twitter accounts and some other web content
- Access: tweet ids, described in [finding aid](#), [blog post](#)

Other Aspects

- Website archiving
- Digital and analog ephemera
- Some outreach
- Iterative assessment of our approach



Take Action Chapel Hill digital flyer

Community-Driven Resources

WITNESS

Yvonne Ng

Documenting the Now

Alexandra Dolan-Mescal

WITNESS trains and equips human rights defenders around the world to use **video and technology** to expose abuses and fight for **justice** and lasting **change**.



- I'm the Senior Archivist at WITNESS, which is a non profit organization that supports people to use video and technology to defend human rights.

- We work internationally with local human rights partners, to support campaigns where there is an identified gap in the use of video, and where we think the use of video can have a human rights impact.



WITNESS's Palika Makam and Jackie Zammuto facilitate video training with Rio Grande Valley Equal Voice Network, Sept 2018. Photo: WITNESS

- It may be easier to explain our work and the ethical approach which informs it with some concrete examples, so I'll talk a bit about one of our programs, namely our US program.
- One of our focus areas in the US is immigration enforcement abuse, and supporting partners who expose abuses by Border Patrol and who advocate for immigrants' rights. In the past year, for example, we partnered with Rio Grande Valley Equal Voice Network to facilitate community trainings on topics like video advocacy strategy and digital security.



CHECKLIST: SHARING VIDEOS OF IMMIGRATION ENFORCEMENT

BE SAFE. BE ETHICAL. BE EFFECTIVE.

Think through the following questions before uploading or publicly sharing footage of someone being questioned or detained by an immigration enforcement agent. Whether you filmed, received or found a video online, it's important to understand that sharing it publicly could put you, the filmer, the victim and/or others at further risk of harm or retraumatization by exposing identity, location, immigration status, etc. Seek advice from a trusted advocacy group or lawyer about how you can protect your community.

If possible, first share your footage with the person or the family of the person who was arrested or harassed by an immigration agent. They should decide how it's used.

WHY AM I SHARING THIS VIDEO?

Be clear about what you want to achieve: Do you want to expose an abuse, counter a problematic narrative, rally

HOW CAN I MAKE MY VIDEO EASIER TO FIND & VERIFY?

Adding visual and written context can help make your footage easier to verify. You can do this by filming landmarks, street signs, stating the time and day, etc. and by providing detailed descriptions after you upload footage.

Keep titles brief and descriptive. Only use factual information to describe the events. Write a description of what happened before, during and after the recording. Always include date, time, city, specific location. Add keywords or "tags" so your video is easier to find.

SHOULD I LIVESTREAM?

Livestreaming can expose people's identities and other sensitive information far more easily than recorded video. It's important to understand the risks before you go live.

Phones can default to sharing your location, be aware of what location details you're sharing.

Describe what is happening and recap what has happened through factual commentary. Work with a partner to

Free resources like this checklist available at library.witness.org

- Our work also includes creating resources like tipsheets, case studies, and videos, like this checklist on how to film immigration enforcement strategically, ethically, and safely. These and many other resources can all be found at library.witness.org. And everything on our library can be downloaded, re-used, and customized by anyone.



WHAT JOURNALISTS SHOULD KNOW BEFORE REPORTING ON BORDER COMMUNITIES

By Michelle Serrano, RGV Equal Voice Network.

All “Eyes on ICE” resources available at: <https://lab.witness.org/projects/eyes-on-ice/>

Search



This article is part of a **WITNESS project**, *EYES on ICE* that offers a deep dive into how to safely, ethically and effectively document immigration abuses through tip sheets, case studies, videos, and more.

Related Posts

WHAT JOURNALISTS SHOULD KNOW BEFORE REPORTING ON BORDER COMMUNITIES

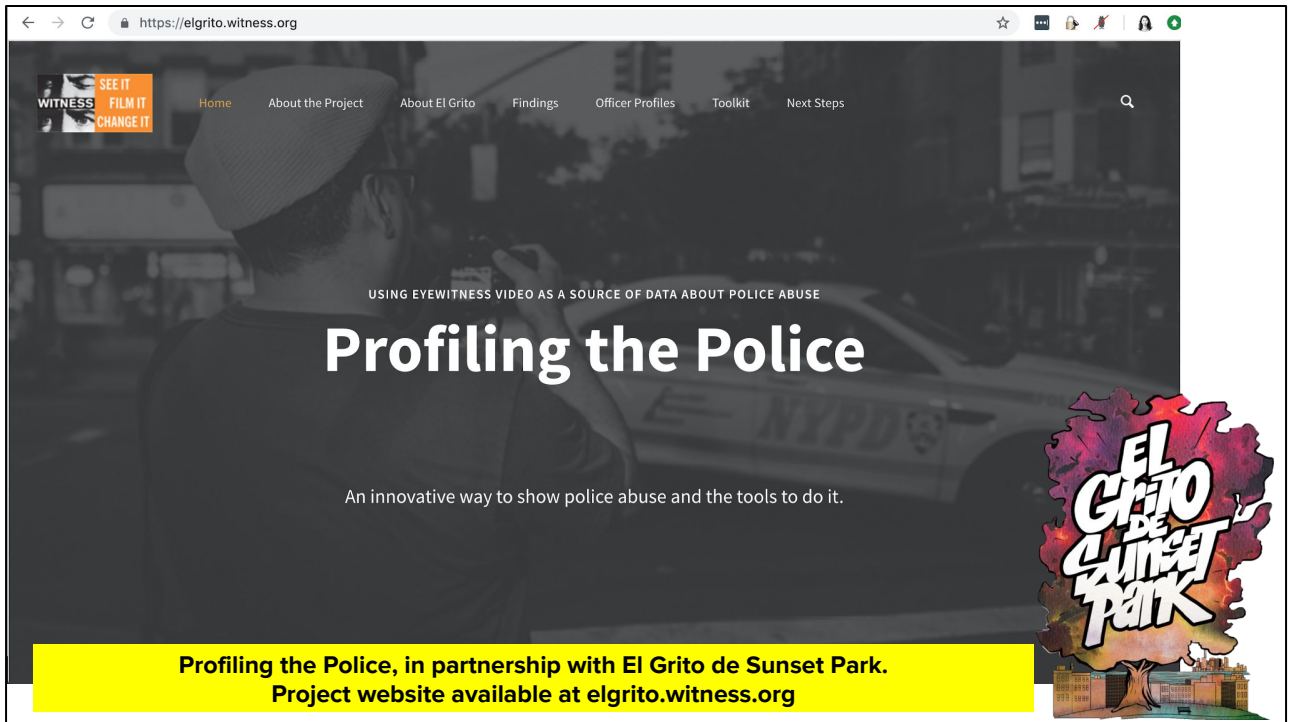
ORGANIZATIONAL DIGITAL SECURITY GUIDE FOR IMMIGRATION JUSTICE

EYES ON THE BORDER TRAINING WITH NEW SANCTUARY COALITION

REPORTING ON IMMIGRATION:

IMMIGRANT COMMUNITIES TO

- I'll also highlight here another resource from our US immigration work that's relevant to our ethics discussion today. It's an article written by Michelle Serrano from the Rio Grande Valley Equal Voice Network entitled "What Journalists Should Know Before Reporting on Border Communities". In it, she gives guidance on responsible and ethical visual representation, use of language, and historical context.



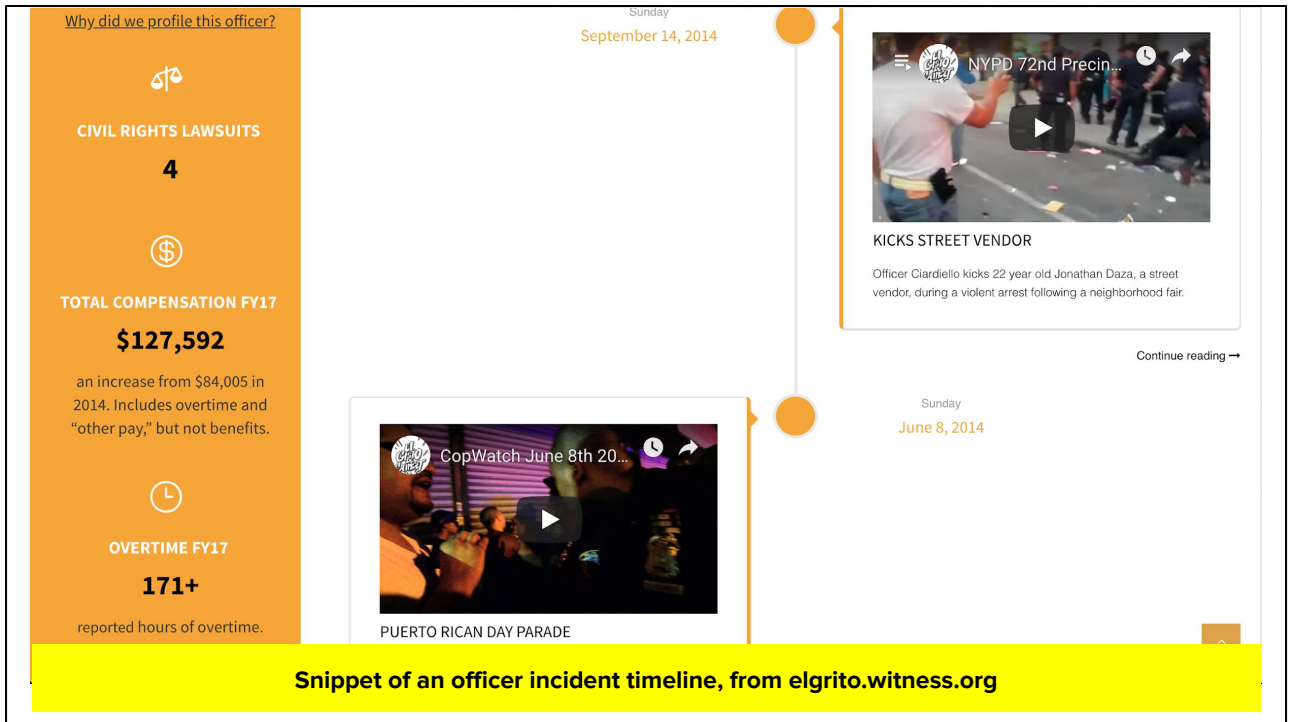
- In terms of archiving, video archiving is one of the areas that we offer training and support on to our partners. That is to say, archiving is meant to help a partner achieve their advocacy goals, not done for its own sake. In the past year, we've worked on two projects in the US with an archival component, both centered on police accountability.

- Profiling the Police was a partnership with El Grito de Sunset Park, a community group in Brooklyn that addresses civil rights and public policy concerns, including policing. Together we created a use case study that explored how copwatching videos and open source or public information could be collected, analyzed, and used to tell stories of police abuse and to address the lack of publicly accessible data about officer misconduct.



WITNESS and El Grito de Sunset Park digitizing mini-DV tapes. *Photo: WITNESS*

- The archival parts of this project included digitizing some of El Grito's older tapes, arranging their digitized and born-digital video files, and creating a metadata scheme for cataloging those videos.




These steps provided the groundwork for the team to the analyze data and create timelines of documented officer misconduct.


<https://elgrito.witness.org/portfolio/elgritocase/>

TOOLKIT OVERVIEW


Throughout this project we documented the tools we used, the processes that worked best for us and the things that made our light bulbs go off. We compiled those learnings into this Toolkit for others to use, share and help grow. The Toolkit can be used as a step-by-step guide, or you can jump right in to the section you're most interested in!




Defining Your Project




Project Planning




Gathering Assets




Preservation




Metadata



Storytelling



Before Sharing



Recommendations

Profiling the Police toolkit, from elgrito.witness.org

- All of the methodologies we used in the project were published in the Toolkit section of the Profiling the Police website.



Working with Berkeley Copwatch on video workflows and database, May 2018. Photo: WITNESS

- We're currently working on a project with Berkeley Copwatch, a community-based organization that has been operating since 1990. We're helping them to streamline their workflows for ingesting videos from copwatching shifts, and helping them build up their internal database of incidents and documentation, so that they can better track their data in support their advocacy efforts, and so they can make it available to other police accountability efforts nationwide.



WITNESS Brazil took part in a Data & Policing Convening, organized by Invisible Institute and WITNESS, Chicago, July 2018.

- These examples are from our US program, but we also do trainings and collaborations like this in Latin America, the Middle East and North Africa, Sub-Saharan Africa, and Southeast Asia. Our global structure allows us to connect across regions to help build solidarity and share learning. So for example, we strengthen our work on police violence in both the US and in Brazil by bringing team members and partners working on this issue together on various occasions to share strategies and methodologies.

GUIA de ARQUIVAMENTO DE VÍDEO para ATIVISTAS



ab
pa
associação brasileira de preservação audiovisual
archiving.witness.org

**Portuguese version of Activists Guide to Archiving Video,
co-produced with Associação Brasileira de Preservação
Audiovisual (ABPA).**

- We also create training resources within regions, that we translate, customize, and re-purpose in others. For example, in 2017, we worked with amazing archivists and volunteers from ABPA - in English, the Brazilian Association of Audiovisual Preservation, to create a Portuguese version of our Activists Guide to Archiving Video, which was originally published in English, Spanish, and Arabic.



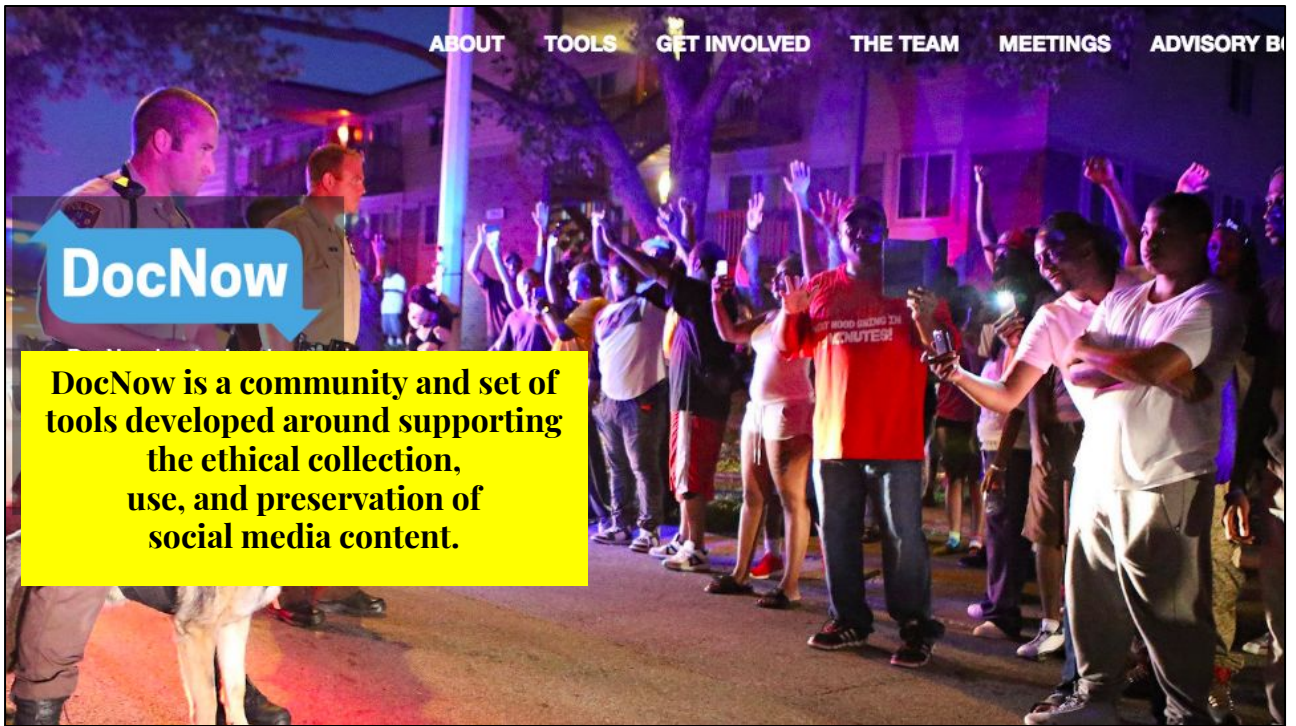
"Arquive!" video (Portuguese) at <https://www.youtube.com/watch?v=ZRIG-Qea8qM>
"Archive!" (English version) at <https://www.youtube.com/watch?v=vUWLqeEf91s>

- Around that time, WITNESS also created a training video that explains the value of archiving for human rights advocacy, produced for use in Brazil, but also with an English version.

Ethics...

- ❑ Non-extractive.
- ❑ Building collaboration and trust. Transparency.
- ❑ Responsibility to share what you know/learn.
- ❑ Honoring others' knowledge and investment. Amplifying others' voices.
- ❑ Informed consent / risk-taking and security.
- ❑ Beyond English, Global North perspectives / resources.
- ❑ ...

- There are a lot of ethical considerations baked into how WITNESS does its work, some of which I've jotted down here, and hopefully we can discuss in the conversation next, but hopefully that gives you an introductory sense of what we do and how we do it.



13 slides, ~30 second a slide



COMMUNITY

THE TEAM



Dr. Meredith Clark
Academic Lead, Co-Principal
Investigator



Bergis Jules
Project Director, Co-Principal
Investigator



Trevor Muñoz
Administrative Lead, Principal
Investigator



Alexandra Dolan-Mescal
UX and Web Designer



Francis Kayiwa
DevOps Engineer



T'Sey-Haye Preaster
Project Coordinator



Ed Summers
Technical Lead



Raffi Vigilanti
Software Developer

Former Team Members

Vernon Mitchell
Administrative Lead 2017
Chris Freeland
Administrative Lead 2016
Timothy Cole
Project Manager 2017
Desiree Smith-Jones
Project Manager 2016
Dan Chudnov
Data Engineer 2016-2017

GET INVOLVED



Mailchimp

Sign up for emails



Twitter

Follow us



Medium

Read and respond



Slack

Join the conversation



GitHub

Participate in our code

ADVISORY BOARD

Advisory board members are active participants in the development of DocNow. Board members were carefully chosen because of their expertise as archivists, librarians, computer scientists, historians, software developers, librarians, sociologists, and journalists, and in many cases their work has directly influenced how the DocNow project was designed. We're extremely thankful to this group and their continued support and active involvement in the project.

Natalie Baur, Preservation Librarian, El Colegio de México

Tressie McMillan Cottom, Assistant Professor of Sociology, Virginia Commonwealth University

Brian Dietz, Digital Program Librarian, North Carolina State University

Jarrett Drake, Digital Archivist, Princeton University

Meredith Evans, Director, Jimmy Carter Presidential Library and Museum

Jonathan Fenderson, Assistant Professor of African & American Studies, Washington University St. Louis

Deen Freelon, Assistant Professor of Communication, American University

Jessica Johnson, Assistant Professor of History, Michigan State University

Robin Katz, Public Services Librarian, University of California at Riverside

David Kim, Project Manager, Colored Conventions, University of Delaware

Mark Anthony Neal, Professor of African American Studies, Duke University

Michael Nelson, Professor of Computer Science, Old Dominion University

Yvonne Ng, Senior Archivist, WITNESS

Matt Phillips, Lead Developer, Library Innovation Lab at Harvard University

Rashawn Ray, Assistant Professor of Sociology, University of Maryland

Nicholas Taylor, Web Archiving Service Manager, Stanford University

Dexter Thomas, Writer, Los Angeles Times

Stacie Williams, Learning Lab Manager, University of Kentucky

Micah Zeller, Copyright and Digital Access Librarian, Washington University St Louis

MEETINGS

Building a community of practice around social media and Web archiving is one of the most significant aspects to the work we are doing in the Documenting the Now project. Mellon has generously funded us to hold two face-to-face meetings, one in 2016 and another in 2017. These meetings allow us to bring our advisory board together and incorporate live streaming to engage with a larger audience.

March 22-24, 2018 Ethics in Archiving the Web Symposium (New York, NY)

December 11-12, 2017 Digital Blackness in the Archive: A Documenting the Now Symposium (Ferguson/St. Louis, MO)

February 15, 2017 #DocNowCommunity and #SNAPRT Twitter Chat: Ethics and Archiving Social Media: Perspectives from Students and New Archives Professionals

August 21-23, 2016 Advisory Board Meeting at Washington University in St. Louis, St. Louis, Missouri

August 11, 2016 #DocNowCommunity Twitter Chat 2: Social Media Research

July 14, 2016 #DocNowCommunity Twitter Chat 1: Project Introduction



TOOLS

Twarc and the Hydrator: Working with Twitter's API

A command line tool to pull down Twitter JSON, and an application to rehydrate IDs

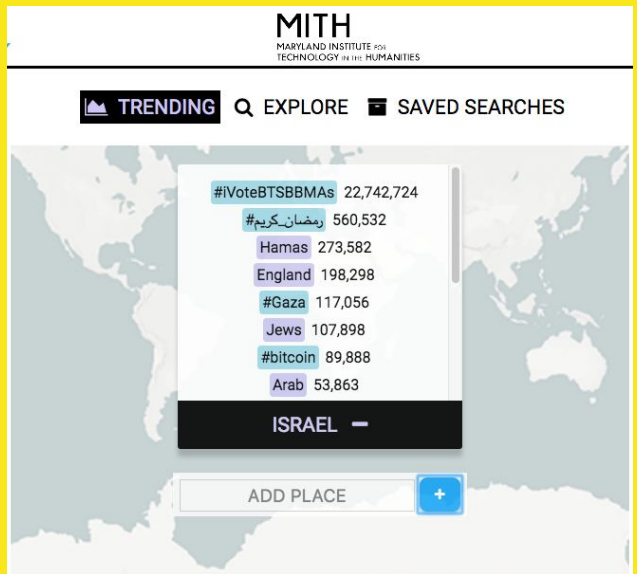
The logo for twarc.py, featuring the text "twarc.py" in a monospaced font inside a light gray square, which is set against a yellow background.

```
twarc.py
```



DocNow: An Appraisal Tool for Twitter

View regional trends, key
players, top media and URLs,
find patterns



The Catalog: Publicly Shared Tweet IDs

A catalog of twitter data - work
in progress!

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 [Iwarc](#) 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.

115th U.S. Congress Tweet Ids

Creator: Justin Littman, Yonah Bromberg Graber
Tweets: 2,041,399
Published: 2019-02-05
Date Coverage: 2017-01-27 - 2019-01-02
Tags: politics, government, Congress

► Description

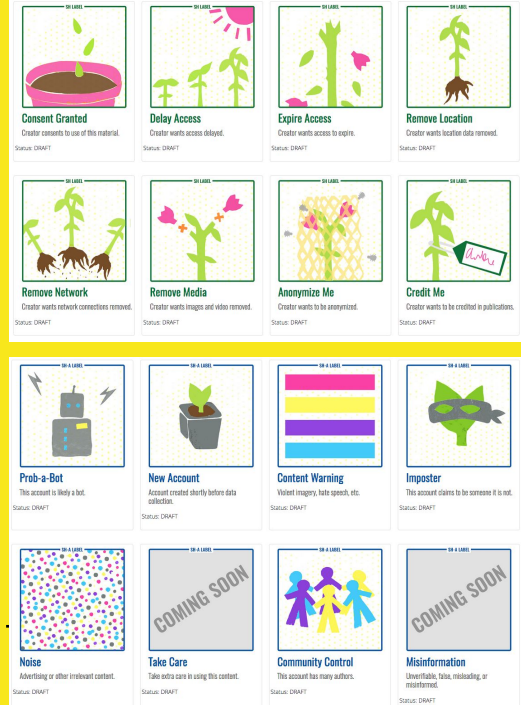
2018 U.S. Congressional Election

Creator: Justin Littman, Laura Wrubel, Dan Kerchner
Tweets: 171,248,476
Published: 2019-02-05
Date Coverage: 2018-01-22 - 2019-01-03
Tags: elections, Congress, politics

Social Humans: Labels for Ethical Reuse

SH Labels empower content creators to indicate if and how they want their data to be used.

SH-A Labels allow archivists and academics to share contextual information



Recommendations

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.

Next Up: Documenting Activism Now



Discussion