Surveillance and Social Media Archiving

On September 27th, 2016 I was part of a panel titled "Concerns for Data Scholarship" at the Library of Congress' symposium, Collections As Data: Stewardship and Use Models to Enhance Access. The other panelists were Nicole Saylor of the American Folklife Center at the Library of Congress and Maciej Ceglowski, creator of Pinboard. Below are my brief remarks. My slides are on google drive. Also check out the full Collections As Data agenda.

At the basic level, Documenting the Now is a project to build free and open source tools that are easy to use, for collecting, analyzing, and sharing twitter data. It's a collaborative project between UC Riverside, the Maryland Institute for Technology in the Humanities at the University of Maryland, and Washington University in St. Louis. The DocNow development team has been hard at work over the past 10 months and I'm really excited for what's to come. Though today I want to focus less on the technical aspects of our work.

What's been the most exciting part of the project in my opinion is how much people from all kinds of backgrounds have engaged with some of the ideas we've been addressing. And I think those ideas address some of the more serious implications for building collections of social media data.

Because our work was inspired by the activism and protests that followed the police killing of Michael Brown in Ferguson, MO, in

2014, I think from the beginning of the project we've felt we had a responsibility to not forget that there are, in fact, people behind all this data. That's why DocNow has such a focus on the ethics of collecting this type of content for long-term preservation. We're really interested in considering how our building of these collections might affect people's lives. It's also why we're being transparent with our work, while trying to help build a community of people who also value these ideas.

So really, DocNow is about a couple of things in my mind, it's about valuing people enough to care about how we collect and steward their data, and it's about helping to build a community of archives professionals and other folks committed to engaging with content owners and creators in equitable and safe ways as we collect their data. Those two things are priorities above our technical work on the project. And a lot of credit really has to go to Ed Summers for the project being framed this way. Ed has been my partner in crime on a lot of this work and he is also one of the principal investigators on DocNow.

We all agree that there is immense value in social media data especially as it relates to our work in archives and libraries. I'm primarily interested in collecting that type of data, especially twitter, because I think it presents tremendous opportunities to document some aspects African American history and culture. For example, according to a PEW Research study, young African Americans use Twitter at a much higher rate than other groups. So in essence, a large number African Americans have found a space where they feel free to share and engage publicly in issues that matter to them, and especially considering how little information we hold in our traditional library and archives collections about African Americans, I think this is a good opportunity to at least learn about those issues, if not to collect data on them.

We've also seen the value platforms such as Twitter have had in amplifying voices in the current movement for black lives, the Arab Spring, and several other social justice events that have played out online. The brilliant work of Deen Freelon, Meredith Clark and Charlton McIlwain showed us how powerful Twitter was is spreading news and narratives during Ferguson. I highly suggest you check out their work if you haven't.

So all of this is good. But we should also acknowledge the significant responsibility and embrace the challenges that come with collecting, preserving for the long term, and making that kind of data accessible. And we should be prepared do this work in ways that don't compromise people's safety, disregard their rights as content owners & creators, or present their data in ways that distort original intent.

We're not the only ones interested in this kind of data. This screenshot is from a report last week by ACLU California that focused on the growth in use of social media collection tools by law enforcement. One of the better-known companies in this area is Geofeedia.

How will our collections of social media data be different than those built by law enforcement and private security companies? How will our tools be different? Here's an image of two prominent activists, Johnetta Elzie and Deray McKesson, who became well known during the Ferguson protests,being labeled as "threat actors" by a private security firm. This is related to the Baltimore Uprising from the spring of 2015 after the killing of Freddie Gray by Baltimore police. So it's a scary situation because these companies are increasingly interested in this data as a way to punish people for being active citizens.

When Ed Summers first published a blog post about the Ferguson twitter data set that we collected during the first month of that event in 2014, a private security firm was one the first groups to reach out to him asking if they could get access to the data. It's a real concern for those in our profession to be aware of. For example, how do we make sure that the massive Twitter data archive being built right here at the Library of Congress doesn't become a tool that these groups can use against already marginalized people who's only request is that police stop killing them?

How will the library respond to requests from private security firms and law enforcement for that data? Part of the answer is that we have to engage directly with the people generating social media data, to understand how our work in collecting this type of data might affect their lives. I think that will be key in helping us develop strategies around the collection, preservation and access to this type of content. It's difficult but it's possible.

And it's especially tough because of the number of people who engage with an issue on a social media platform at any given time can be daunting. I'm not sure how large the LC twitter dataset is but I bet it's significant. In the past week alone Ed Summers has collected almost two million tweets related to the police killings of Keith Scott in Charlotte, North Carolina and Terrance Crutcher in Tulsa, Oklahoma. But there are ways to engage if we're willing to drop some of our traditional models of building collections that prioritize our ideas about professionalism and the myth of neutrality over the wishes people and communities.

Last month the DocNow project hosted our first advisory board meeting in St. Louis. This was an opportunity to get our awesome board members, and friends of the project together for a deep dive into many of the issues we've been raising around social media, web archiving, and ethics over the past year.

We had six really great panels of insightful and challenging discussions. So check them out on our website at docnow.io if you get a chance. But I'm going to focus on one of those panels for the last few minutes because I think it's a great example of the type of community work we're going to need to engage in in the future if we want to building social media data collections as archives in safe and equitable ways.

The panel was made up of a four activists who we some of the many organizers who led protests in Ferguson after Michael Brown's killing. It included Alexis Templeton, Rasheen Aldridge, Kayla Reed and Reuben Riggs, and Dr. Jonathan Fenderson expertly moderated it. I'm really thankful to the activists for joining our meeting because they added a richness and realness that we otherwise wouldn't have had, as they shared stories about their lives before, during and after those pivotal days in Ferguson.

So I'll stop here and let you hear a bit from them about their experience. In this clip Alexis and Kayla are responding to a question about how they want the movement to be remembered. Their comments span from 50:17 to 56:12 in the video. But I highly recommend you watch the entire hour and half conversation as well and all the other panels from the advisory board meeting.