

The Stanford Open Policing Project: An Interview with Professor Cheryl Phillips

Camryn Pak
Stanford University



Cheryl Phillips is the co-founder of the Stanford Open Policing Project, which is a cross-departmental effort to collect policing data to evaluate racial disparities within the system. She has been teaching journalism at Stanford since 2014. Phillips is also a founding member of the California Civic Data Coalition, a group that aims to make California

campaign finance data accessible, and most recently, she founded Big Local News. Before coming to Stanford, Phillips worked at The Seattle Times for 12 years as both an investigative reporter and editor. She was involved in the Pulitzer Prize-winning coverage of a landslide that killed 43 people in 2014. She also was the sole editor in the newsroom when four police officers were shot at a coffee shop, and she was involved in coverage of the event that won a Pulitzer Prize in 2009. Additionally, Phillips has twice been on teams that were Pulitzer finalists. She also served on the board of Investigative Reporters and Editors for 10 years, and she is a former board president. Phillips also worked at USA Today and newspapers in Michigan, Montana, and Texas.

Pak: Can you tell us about how the idea for the Stanford Open Policing Project came about?

Phillips: Sure! So, I came to Stanford in 2014, and I taught my first data journalism class in the winter of 2015. For that class, I assigned students a public records request assignment so that they had to pursue some data. When I worked at the Seattle Times as a data journalist and investigative journalist, I had early on, maybe in 2002 or 2003, done a story about state patrol data and whether racial profiling was happening. And so, I had already collected data and analyzed data and done a story in Washington State, so I knew that some state patrols had this data and wanted to see how many. So, I thought this would be a good assignment for the students they would learn how to negotiate for records.

And it also tested out something for me, because I had come to Stanford with this idea of something called Big Local. It's now officially called Big Local News, which is the idea that a lot of data that journalists use is collected used for a story and then forgotten. But the data could be collected and aggregated and normalized and used for much greater impact if journalists had a way of sharing it effectively. So, I wanted to see if it was even possible to collect similar data from very disparate sources—many states, cities, counties, and so on—and bring it together and use it.

So, the Open Policing Project was kind of my test, in some ways, a pilot project which took on a life of its own. What happened was the students in that class came back with a lot of data, and we ended up with millions of records. And just through the course of life at Stanford, I met Professor Sharad Goel, who's in the School of Engineering, and he has done a lot of work looking at disparities and bias and discrimination. Professor Jay Hamilton introduced us, and when we met and started talking over coffee, I started talking about the data my students had collected, and his eyes kind of lit up because nobody had collected that data and brought it to one spot before.

He had data scientists and Ph.D. students eager to look at that data to see if they could do some more analysis of it for disparities and bias. We decided to partner, and my students, and some students at the University of Maryland in some media law classes pursued more data. We went from state patrols to looking at the largest city in every state, and then to just other large cities that made sense to collect. Now, we have over 250 million records from 33 states and 57 cities.

Pak: Wow, that's amazing! As a collaboration between the Stanford Computational Journalism Lab and the Stanford Computational Policy Lab, the project involves individuals with different backgrounds and specializations. What kinds of people are involved in the project, and what roles had to be filled to make this project a reality?

Phillips: Yeah, we really do have a mix. We've had data scientists, engineers, journalists—both those who don't necessarily know how to analyze data and data journalists—and teachers because we also trained journalists on how to analyze this data and how to tell stories from this data. So, we did a whole range of things, and we needed everyone. We couldn't have developed the very robust statistical measures to evaluate whether discrimination was happening without the data scientists. And, you know, Professor Goel is very well-steeped in the tests that are used to look for discrimination and bias, and so he was able to take the analysis to a different level.

For example, we ran a test called the “veil of darkness” test in a way that really hadn't been done before because we had so much data that was national. I think that really moved the data science forward and what had been done with looking for bias and discrimination forward. I think that work is really important. But the journalists who in the public sphere, took that test, The Los Angeles Times, for example, and did their own data analysis. They then added to that the reporting that they needed to be able to effectively tell the story. That happened in a lot of places, and so we needed both of those things—the expertise in both of those areas.

Pak: Did this year's events involving police brutality and racial injustice impact the project in any particular way? Were there any changes that were made to the project as a result of these events?

Phillips: No, I don't think so because we'd already been kind of working away on it. I think we got a lot more interest in the project, so we've done more interviews for talking to more people about the data. We've had more people offer to share the data they've collected with us, which is fantastic. So, from that perspective, not really, except that I think it just gave us more of a platform to get this reporting out there and to help journalists tell stories from it.

Newsday just did a story, for example, in Suffolk County where they kind of went through the same steps. So, people are still using this data and comparing it to the data that they collect in their own jurisdictions and areas, and I think that that's really important. But I will say that we have become in the last year, before George Floyd, very involved with an effort called the California Reporting Project, which is 40 newsrooms across the state that have been collecting police disciplinary records ever since there was a law passed to make those open. We're working with them to create a structured database of all this police disciplinary information. Eventually, some portions of it will become public. But it's really, again, the idea that if we can pull all of this into one place and share it more effectively with journalists and also to the public and to advocates and criminal justice experts, we're going to see more work done in this field.

Pak: Finally, can you tell us a little bit about Big Local News and any other initiatives that you are currently part of?

Phillips: Sure, so Big Local News is basically the outgrowth of the Stanford Open Policing Project. The Stanford Open Policing Project showed that there was one kind of data that we could collect, but there are other areas where we could collect and disseminate really vital data. So, we decided to try to build a platform that would allow journalists to easily share information, and we created Big Local News. And so, a journalist can go in, create a project, upload their data to that project, add other journals that they want to share it with, and then keep the project private.

Then, upon publication, they can make that project open if they want and share the data with anybody using the platform. And then, if they're interested, we can also archive that data with the Stanford Digital Repository, which means it's saved forever. It also returns back a URL, called a persistent URL, so that journalists who have the archived data can use it and put it on their website to share the data with readers in the interest of credibility and transparency.

This platform launched officially in March, and something else happened in the U.S. in March—the pandemic. When we launched the platform, we had a handful of users, and within weeks, we had hundreds of users because we started sharing COVID-related data. Now, we have over 1,3000 users, and we share a couple of dozen data sets related to COVID, everything from case counts to hospitalizations to hospital beds to nursing home information to evictions data. We have almost 1,400 users and 155 projects. I think it has become a really useful resource for journalists.

The other thing that we did was partner with the Google News Initiative and created a COVID case mapper. One of the things that journalists were trying to do was to tell their readers what was happening with case counts in their area. We have some great national maps from The Washington Post and The New York Times, but it's much harder to build your own local map. So, we built a map that allows anybody to take a county or state view, embed it on their site, and share it out. And so, we've gotten hundreds of thousands of page views and tens of thousands of users actually embedding it on their sites. And I think that saves them from having to build that themselves. So, I think that that's an example of collaborating in a way that provides something of use to journalists so that they can spend their time on local accountability.