DATA SCIENCE ETHNOGRAPHIC INTERVIEW

By Mya Strayer, Oct. 27 2025

<u>Interview Transcript:</u>

[DISCLAIMER: I intentionally chose someone who wasn't a data scientist by title, but rather worked at a small enough company that outputs data-driven work. I know this person rather well (and I personally know they've done the work of a data scientist) and was hoping to highlight in this piece how broad the work of 'gleaming insights from data' has become.]

So I'd like to start with just you describing what fire.org does and then specifically what your role at fire.org is, and if you could talk about that and kind of connect it back to, like, how your role specifically and your company generally uses data as part of its workflow process.

It's a nonpartisan, nonprofit First Amendment advocacy organization. We work with campuses and in society at large across the country to not just defend the First Amendment rights of people who have been censored or facing some kind of repercussion for their speech, but we also work to educate people on what their rights are and to build a cultural acceptance of free speech principles across the nation. As far as my specific role, my title is student press Program officer. So what that means is that I am in charge of all of our student press freedom initiatives.

I work with student journalists who are writing for all sorts of different outlets or including some who do broadcast or radio journalism and provide them with legal resources both before they run into issues so that they feel confident hitting publish and

then connecting them with legal resources that they might need if somebody gets upset about the coverage they produce.

As part of that role, I work with a team whose whole job is to do First Amendment advocacy on college campuses. And that team relies extensively on data about our cases, about our interventions, about the way that you to make sure that every time we spend an hour something that hour is being spent in a way that either boosts people's understanding of the First Amendment or leads to successful results when we're defending a particular student faculty member, oh, you know, other community member who is facing some kind of censorship.

Perfect, thank you. So then I would like you to think about a specific data science project that you have either worked on recently or been a large part of. And if you could just tell me a little bit more about the general description of what that project was, where the data was coming from, slash, how it was going to get used, and what that goal was behind that project.

A couple of months ago, my team, which is relatively small for the amount of work that we do, we, I think, have eight people on the team. We did two things. One, we moved our CRM, which is, you know, we just use CRM for case management. We moved that from a platform called Streak to Salesforce. And with the move, we realized that we had a lot more data at our disposal. Because Salesforce allows for better data collection and better data presentation. And we wanted to find a way to use that as accurately as possible to measure workflows. And so that when we were assigning cases before, we just had a round robin system where it was say, a press case, in which case I got it. It was first, you know, this person gets it, then person number two, then three, then back to number one. That led to all kinds of problems because there were some people who would get five cases and none of them were urgent. And so they still had room on their plate. But then someone might get a case that required 15 hours of work and they would get assigned

another case just as quickly as the last person. So we wanted to find a way to take the qualitative experience load and turn it into something that program associates could use when they were assigning cases to make sure that they were assigning to someone who actually had availability.

What were the major flags that you were looking for within the data to help indicate urgency or help indicate where, I guess, the amount of work that it should be needed for each case.

Yeah, so the first, I mean, the first thing that we were looking at was we really didn't have an actual representation. I mean, you could look at the list of cases in the old system and see this person has this many cases, but that was actually not a, that was not a datum anywhere. It was, you know, there were data on however many cases, but there was not a datum of it. This person is working on x many cases.

So that was that Step one was just how many cases are working on. Step two was where in the process is that case. The system that- the workflow that we've used for a very long time has multiple stages that a case can go through. And there are some stages where cases are very active. So that's when we're actively drafting some kind of intervention or when we are discussing a case in a meeting or doing the initial intake review. Like those are a little more involved. We also have cases where we've already intervened or just waiting for something, or all we're doing is trying to do information gathering up front. And those cases aren't as urgent.

And the program administrator can see of the, you know, 17 cases that I'm working on right now, where are, how many are in this stage? How many are in that stage? And then the last thing is just it was difficult in the system to replace the qualitative experience of workload and how busy a person felt, how much time a person felt they had. In part

because yes, we can get pretty granular about how much work a person is doing on cases, but all of us have aspects of our job, not casework.

In addition to doing casework for student journalists, I'm going to conferences across the country. I'm presenting as we speak in Dallas airport because I'm going to speak to a group of faculty members. So today, even though my caseload is actually pretty good, it's high. But one would not look at my schedule and think, oh, he cannot take another case. But because I'm doing this other thing, we still needed something to capture and so what we ended up doing was creating a color system. Essentially, you know, the 'traffic light of green' means I have nothing but time. Please give me more work, all the way down to purple, which means I literally cannot check my email. Please do not send me anything.

Was this change towards doing case management in a more data-driven way, was this because of the affordances that you guys found within Salesforce or was this more of a company-wide overarching cultural vibe of needing to have more data driven workflows or reasoning behind your case assignments?

Yeah, it was a combination of both. I think Salesforce was a tool that enabled us to pretty easily use data that we already had to implement some kind of a system. But we had a lot of ongoing issues with workload over the course of really– I don't remember which semester it was at this point, but there was one semester where we just had a ton of issues with workload because our team was understaffed. And so we knew we needed to find a way to better assign cases to the right people.

In addition to that, we wanted to have data not just for us when we were trying to decide who should do X, Y or Z internally, but also that we could bring to higher ups in the

organization to say, look, this is the workload that we collectively have and this is how much each person is doing. We need another person or to better allocate, we need to reassign this person from the, you know, 50% they split with this other department.

Could you briefly describe what your role was and when your role actually started or where in the project were you pulled in? And the roles of the people who were managing this project, managing this- this migration. And then where in this like the overarching project did you get brought in and what was your specific role as part of this migration?

The Salesforce migration was a joint project between us, our driving team, which was already on Salesforce, our public advocacy team, our litigation team. Then finally we had basically two people who were, who had become our data team. And they were the ones kind of spearheading the migration. They added a third person later on and the way that that team was structured was they had one person who was the manager of our data system. So that is yes, the Salesforce migration, but also some of the other departments use other communication platforms. The development team obviously has their own database primarily through source, but they also have their own separate like prospecting databases that they use. Those are also managed by that team.

But while there's one person who's in charge of the entire data project of fire, there are also two people underneath him. And each of them has a set of departments that they work with. So like my team has one person who is dedicated to us and who anytime we need something we can go to him. And then the other half of the organization has another person who's dedicated to their projects. Those departmental splits mostly align in terms of kind of mission areas. So, like, the development team and our administrative team are managed by the same person, but our campus team and our litigation team are not. And

that's because we have a disproportionate amount of work supporting programming, so case oriented departments compared to administrative departments.

So then just to clarify, you had these more specific teams that were almost more like the user experts or system experts of specific parts, like your litigation or your administration team that is an expert on just that part of this, that is then communicating and conversing with your data team to make sure that, like, the data structures themselves can replicate what the workflow needs to be done by the users.

Right. And my role was a little– Well, I guess you can't have grades of uniqueness, but my role was unique in that I was one of the people who was a workflow expert for the campus team. But I also had pretty extensive experience working with the litigation team, with the advocacy team. And so I was sort of a second person who provided an overarching view of what the user experience was like and what the various workflows were. So a lot of my job was when somebody would say, for instance, when someone who was currently on the litigation team and say, hey, we want to do this, I would step in and say, but the campus team uses that thing this way. We use that set of data in this way, and vice versa, that when someone from the campus team would say, we want X, Y or Z, I would say, well, the reason it's done, the reason we have A, B and C instead, is that litigation needs that. And they structure their data slightly differently. And we've just sort of shoehorned ours into that system.

Could you talk to me a little bit more about where, or when your team was coming in and when you specifically were coming in, where was the project already at? Where was the data system or more so its structure at, how far along was that? And where was your expertise really being brought into?

Yeah, I- I think what I did at each stage of the project, I think varied. But I think I was, you know, on the broader data migration side from the very beginning. From the minute we decided that we needed a different system to manage the case data. I was a part of that project. I was talking with the data people, I was talking with the executive team and I was talking with our departmental managers. And my role changed within the organization to be fully housing. Because when this project first started, I was still an executive assistant. And then eventually as we continued to build out the Salesforce instance, I moved to the campus team and took on my current role. So I went from a very broad generalist to somebody who had expertise, but ended up bringing that broader expertise to the project all the same.

But that was only after I only moved to the campus side. After we had Salesforce. We had done all of the migration and at that point we were just refining and coming up with new ways to use the data.

So this actually leads really well into the next major section of this interview, which would be the project stages and activities. So you've already addressed that there were these different stages. It sounds like there was a big stage of an initial mass migration.So in some way like the actual "We're trying to replicate the old system" and that there then was a later phase of now that we have this data, we can ask what we can do with it. So, if you could talk more about maybe those two specific phases, and what the transition between them looked like?

Yeah, so the first six to eight weeks were just actually setting up the new system and making sure that we were replicating as much as possible from the old system, building the workflows that people were used to back in [the old system]. And while I was involved with telling them what the user experience was like, I was not under the hood much with that. I think then there were probably two or three weeks troubleshooting where it was

just trying to restore functionality that we didn't realize had been lost, replicate functionality that didn't natively exist in Salesforce's CRM. Restoring data like copying and just, you know, copying and pasting data that didn't quite make the transition correctly from the old system, but where the transition happened from that establishment. And I'll combine those into one phase, right?

The establishment and troubleshooting phase to a phase of exploration and experimentation was when each team had a problem. And it wasn't usually a data problem, it was usually a process problem. And they would realize that the data we now had could be better leveraged to solve that problem. So the workflow example that we talked about at the beginning is one of them. Another one that affects me personally, one of the things I do beyond the student press work is I am in charge of our legal referrals network, which is just a list of lawyers, their addresses, their emails, their phone numbers, and what they want to work on.

And as part of that, it used to be that I had to go into a spreadsheet, look through the spreadsheet at where every lawyer was based, and decide if that was close enough to where the case was to include that lawyer in our generic referral.That didn't take a long time. And I was good enough at geography to not have to look much up, but it was definitely more cumbersome than it needed to be. And when we moved that spreadsheet into Salesforce as we were doing that migration, I said, you know, is there a way that we can take the location data that we now have stored in a way that the computer can understand and not just in a spreadsheet and automatically generate that list of people?

And so eventually we worked with the data team and we developed that. But I think that's a good example of how these other projects got started. Was that people would say, hey, there's this process that's cumbersome. Can we fix it? And there was a way. I will say

while I don't think it was: people saw the data that were in Salesforce and said, oh, we can use that data to do X or those data to do X, Y or Z. I think people were more open to consider data driven solutions to problems because they saw the wealth of data that we were now storing that we didn't have before.

Could you talk more about what that specific maybe week or two of what that process was like from the finding of the problem, like this is my idea of how we could do this, to bringing that to the data team, and then to actually being able to see that within your guys's Salesforce instance and utilize it?

Yeah. So the idea started because we were migrating to a spreadsheet and the data team had asked what I was using the spreadsheet for. And I said, well, you know, I go in and I find the address and that's that. And try to get as close as possible or within, you know, a certain radius of the place where the violation happened or where the litigation would ensue. And as we were talking about it, the data team asked how often these referrals were on campuses. And I said, well, you know, they're fairly frequent, probably 50, 50 on campus versus off. And they said okay. So they built out a couple of different automations to pull contact data from our web form into Salesforce and also from the old spreadsheet.

And once that was there, they asked me basically like, how cumbersome is the process of going through your strategy. And I said it's not bad, but you know, if there's a way that we can automate it. They said, well, we have all this iPad data from the government about where these various schools are. Lots of other things about the school too, but specifically they had geographic data and they said we can just take that, build a radius around it and then have Salesforce include the attorneys who were in that radius. And I said that sounds lovely, let's do that. And so they built it out. And eventually we expanded it so that for cases that didn't involve universities. And this was probably– So I guess to fill in the

timeline, it probably took about a week and a half from that initial suggestion of theirs to when it was fully implemented on our end.

And there really wasn't much tweaking that needed to be done with it. It just worked off the bat. It did exactly what I needed it to. So then a couple weeks later, I ran into a case that was not from a university, and so there wasn't the iPad's data to tell Salesforce where the case was. And I said, well, is there a way that I can just put in lat long data and have it do the same thing? And then the only step I have to do is manually input where this case is occurring. And they said, yeah, that should be pretty easy. So they built it out for one department and I said, great, can you add to it? And then a week later I said, great, can you add it to this other department that doesn't use iPad data? So by two months later, it was. The system worked for every single department.

And then at one point in one of our departmental meetings, my boss had said, why are all of these cases in referral? And I said, all that's where they need to be if we want to use the referral tool. And she said, well, can we not have it do that? So I talked to the data team and I said, hey, can you edit this too? And I listed the specifics. It doesn't need to be everywhere and it doesn't need to be dominant because I'm the only person who's doing this. So it shouldn't disrupt anyone else's workflow or it shouldn't put buttons where they're not expecting them.

But there are some other stages other than just referral where we might want to find lawyers. And the data team added the button they edited on all the other phases that aren't referral. They edited it underneath a little drop down that I knew where it was and I could go do it, but nobody else would accidentally hit the button and create things that they weren't supposed to. Yeah, yeah, that actually like a week.

You had mentioned about the two to three weeks that you spent of troubleshooting, and I was wondering how in those phases, especially when you have like multiple leading experts or multiple subject experts, what did you guys do to kind of help generalize those different workflows or did you have any issues with, like, I do my workflow this way versus this person does their workflow this way, and now we have different uses for the same data. You kind of mentioned that, that there were some issues with multiple teams wanting to use this data, but for different things, or both use a structure for different things.

What they ended up doing and this, they did this before the migration actually happened. But they created a data governance committee for the organization and that's for the most part, the lower level administrators of each department, because those are the people who are actually manipulating data the most often. And there's somebody from each department, there are alternates for weeks where the primary can't attend, so that there's always somebody familiar with the workflow of their team.

At this meeting, when we discuss something, sometimes it's the data team who wants to make a change and they want to understand how people are using the data before they start making the change. Sometimes it's a department that's looking to do something or a department is having problems with something, and we collectively discuss it to figure out how, how we can harmonize the uses, or in the event that there's no way to harmonize them, which team needs the data to be structured the way that it is and how the other team can adapt to that.

Within the specific steps of the project, was there anything that was particularly easy for you, and could you maybe describe why that specific feature was easy for you?

I mean, I think the easiest thing for me to do was to identify why something was set up, the way that it was set up, or why it needed to be set up a certain way, if it wasn't set up that way already. And I think that that was particularly easy for me one because it fit pretty well with the, you know, kind of day to day work that I do, where I'm, you know, constantly looking for here is my particular fact pattern, what are other analogous things that exist elsewhere, and how do I apply them? And so when we would talk about, you know, this is a field on a contact record, it was pretty easy for me to look at that and say, Oh, this contact, this field is for, you know, the development team is using it this way, and that's why we're talking about it. But I can see how this could be helpful for me as a person managing the network employers. Or I can see why changing this to suit development needs better would make it difficult to reach out to a client, for instance, those you know, those aren't specific examples, but that was the type of thinking that I think helped navigate me throughout the process, and it was made easier by the fact that I knew how a solid two thirds of the organization operated it and what they were doing with the data that they had.

Do you think this– and this is more of just my own curiosity— do you think that part of that was due to the nature of the role that you play or due to the person that you are? I'm wondering how much would you ascribe to you as a person and like versus how much do you think it was because of the ease of the system?

I think it was mostly individual factors, especially where it was identifying a use case, beyond how it was presented in Salesforce, because of the way our instance is, some of the time we would be talking about a field that exists on every contact but is only ever used on one set of contacts, and that doesn't factor into the workflow for the third set of contacts at all, which I think made it difficult to see. And there are also some cases where, for instance, I need my attorney network to be set up or to work the way that it's

supposed to. I need the data presented in a certain way, but the development team might want to just know that that person is involved with the legal network when they want to interact with them, right? They probably do need to mark the data differently than I do, but that leads to complexity, and it leads to a system that is less easy to understand when you're just an outsider looking in and not involved with either of the two sides of it, and it's difficult to because there are these two different ways of presenting the same thing. It's difficult for somebody in the third department to look at and say, Oh, I can use one way or another for this other thing that I'm doing. It's easier than it was, so I'm not going to knock it all together, but I think it is not something that would be intuitive to the average person without some kind of specialized knowledge of the way that the departments work.

I was wondering what the maintenance phase looks like? Are there remaining initiatives to make sure that these workflows continue to work properly, or to continue to explore and readjust those dashboards, or readjust the ways that you can use your data?

In terms of exploration, that's largely being led at the department level. There is some of that that is driven by the data team, but most of that is departments saying, hey, we need this other thing. And the data team is really gracious, and we're supposed to do that. There's also the part of it that, I mean, some of the systems that they have built just break from time to time, because systems do that. So that tends to be more data team driven. The one spot where there is some overlap is in those data governance committees, because it might be that one department went to them and said, Hey, we need a dashboard to manage workload. But in those meetings, what we're talking about is what team A is doing to get the workload dashboard. I think the data team is good at identifying possible new spaces for other departments. I mean, this is how one team uses it, but you could use it to track the, you know, urgency of a particular matter, or you could use it to track the

overall case, or, you know, whatever the specific instance that might work for another department may be, but the data team has been very good about identifying those where they think they exist, and also identifying new features that are being added to the platform that might be useful to various teams.

How does this project compare and contrast to other, we'll say, data driven work that you or your team has done, or even your organization has done. In which ways was this kind of implementation, in which ways was this very similar to things that you guys have previously done in the past, and in which ways was this very different from other types of work that you've done in this space?

It was common in that we had always collected data for our what I'll call quasi external presentation. So presentation to people who cared about our successes as an organization, people who cared about the organization itself. So you know, thinking donors, board members, people who were just invested in the continued success of fire as a corporation, and where it is different than some of the other data we collect is that we also have a team that does research, and they do a lot of data science process survey for hundreds of universities across the country and their speech climate that research in that data science is being done with a purely external audience. Certainly we use it internally. Certainly it's integrated into our network. So for instance, when our policy team looks at University's policies, they'll publish what we call the spotlight rating, which just says, This is good for free speech, this is bad for free speech. This is kind of in the middle. And those data are contained internally. So when I go to write to a school, I can see how their policies shape up, but the primary goal of that information is for public dissemination, whereas with this project, I think pretty much all of it is for either a purely internal audience, like the individual staff members workload, or a quasi internal arguments, which is where you know where the performance data is in essence.

Okay, perfect. Could you tell me about an aspect of the project that you found particularly enjoyable and why it was enjoyable, and if you could talk specifically about, like, how did the Salesforce implementation help make your day to day, data driven work more enjoyable.

I think the most enjoyable part of the project for me was being able to see our workflow as a team, and the workload of my colleagues improved, or, you know, decrease, and the, I think, really, the quality of our work and the timeliness of our work improved, because we also, of course, track how quickly we're getting these interventions out. And it became clear that we were actually getting the interventions out faster because we were assigning cases better.

And that was enjoyable because, I mean, it's a great thing to give to anybody who was passionate about the organization. But after we implemented that, we had other departments coming and saying, like you guys have become more efficient. How can we do the same and that, I think that ripple effect was probably the most important part was being able to then talk to other teams and other administrators and say, This is how we pulled that off, yeah. And this is how you have the same information that we did, like you have the data that we used to pull this off. You also have, here's how you can use it. That's really great.

The last question from the more official side, is just, if you had to do a data science project like this again, is there anything that you would do differently? And why would you do so differently?

I would message the applications of the better organized data sooner and better. I think we did a really good job of capitalizing on the data, the new data that we got, and the better organization of the data we already had. But I think at the beginning there was a lot of distrust, because, to be honest, there are a lot of things that we have to do now. Our improvements to the process, I think we're more efficient overall now. And everything I said that about the greatest thing this project does stands true, but we did make a lot of trade offs in making that transition, especially on the CRM side, and I think there would have been a lot greater buy in organizationally, if we could have articulated up on now, this is what you're going to be able to do. We're going to better be able to better manage your workload, make you more efficient and make your life easier after we make the switch, even if you now have to click maybe more buttons every time you want to put an email in the CRM, yeah.

Coming from a non technical perspective, do you feel as though coming into this project, did you feel like you had a strong enough intuitive understanding of how this data would be stored, or was the initial learning process of figuring out how data gets stored and how to work with the data team strenuous?

I think there would have been a learning curve if I needed to actually understand how the data team was thinking about some of this. I really think that they did a very good job of translating like, I still don't really understand the back end of how Salesforce works. I understand the data that we have. I understand the data that we store. I understand what data point is linked to what kind of record. So I understand, you know, this sort of data is linked to a tech, this sort of data are these sorts of data are linked to a case, these sorts of data are linked to an institution, and these are the types of things we put there. There had to be a lot of conversation to make sure that we were all speaking the same language about what an institution is, is it a foundation, or is it a university, or is it both of those

things? Is it a police department? So there was a little bit of making sure that we all had common language where the Salesforce definition and the colloquial definition didn't align. But I think in terms of the really technical side of it, it's what kinds of databases are being used? How is it actually pulling from the database? Is this piece of information about a contact record being stored in the same digital place as this piece of information on an account record, that sort of thing. I mean, I don't think we've hit a need to really understand, because the data team has been very good about just providing answers in plain English. Or it's only the jargon the user sees, not the jargon that they're thinking of.

Yeah, interesting. And then my last question is, I was wondering if there were any instances within this project where you guys had concerns around the sensitivity of your data, or needing some sort of level of protection of data?

Yeah, that's a really good question, and those conversations I think were fairly easy because the people with the permissions were actually, well— legacy users of Salesforce already had the proper firewalls built around their data. Even the development team, you know, walled off specific donor information, like how much they're giving us. And that was that way, long before everybody else joined Salesforce. So the discussion was really, how do we preserve the confidentiality that we've already created, even as more and more people are using it? There was some discussion about case data and who should be able to see the details about a particular case. But for the most part, the case data, the real concern wasn't making sure that it was secure internally, it was that it was secure from outside people we didn't want to, you know, disrupt attorney client privilege or anything like that. Yeah, but we didn't have much of a problem with most people who work internally seeing it, and in fact, most of what we found was that there were some firewalls that existed, that existed for historical or practical reasons that really didn't have anything to do with the sensitivity of the data themselves.

Okay, perfect. So I'm going to end the recording, but I do want you before I do, could you give me an estimate of how many employees are at your company?

We have roughly 120 full time employees. The project I would say had somewhere between 60 and 70 Salesforce users.

<u>Additional Questions I Asked:</u>

- [I asked quite a few ad-hoc follow-up questions throughout the interview, but I'll talk specifically about the two I couldn't end the interview without asking:]
- *Coming from a non technical perspective, do you feel as though coming into this project, did you feel like you had a strong enough intuitive understanding of how this data would be stored, or was the initial learning process of figuring out how data gets stored and how to work with the data team strenuous?*
    - WHY I ASKED THIS: I asked this question because I wanted to hear more about what made a data team un/successful by their business counterparts standards, and how strongly my interviewee would assess his own knowledge of the data structures. From his previous responses, he seemed to be very knowledgeable about what data they had and how it was being stored, but when asked this directly, he accredited the data team for being effective at 'speaking the same language.'
- *Yeah, interesting. And then my last question is, I was wondering if there were any instances within this project where you guys had concerns around the sensitivity of your data, or needing some sort of level of protection of data?*
    - WHY I ASKED THIS: In the spirit of our last assignment, I wanted to bring some awareness to how differently the word 'sensitive' can be interpreted when

referring to data. Of course, given the nature of his organization, a lot of this data could be defined as sensitive– in the meaning that it could be personally identifiable and could cause harm to an individual's life if the wrong person got access to it– but it could also be interpreted (as it was in his response) as whether or not there would be legal issues surrounding the access to data.

The Salesforce Implementation Process:

In order to implement a new CRM (Customer Relationship Management, I'm not Zork) tool in a nonprofit organization, you'll need to do two things: learn the affordances and structure of the new tools, and figure out how you can most effectively use it for your longstanding processes. While there will be a data team, it's better for as many people as possible to learn as much as they can about the different data structures as they can then use the system as efficiently as possible. The project will start out with a few people becoming experts in the nature of the system and its encodings (the data team), and these people will then meet with the various department teams and attempt to replicate the previous workflow into the new software. If there are any issues with multiple teams needing different things to be encoded by the same data structure, there will be a prioritization process that determines which encoding would be the most generalizable. This work is conducted by the data governance teams, which includes the people who use the data most often and they report their decisions to their general team. Then, once all of the current processes work relatively smoothly, it will be the responsibility of the department-wide teams to either find bugs or optimization opportunities. Either of which can be brought to the data team, who will then work with the respective data governance teams to implement the new functionality or change the existing. Finding ways in which you can better use the data at your disposal can have a positive impact on your career, either through a direct promotion or increased responsibility in your work. Your ability to think in terms of data optimization can also inspire others around you to do the same, making your overall workplace more efficient and data-driven. Maybe everyone's a data scientist (except Zork, he's my friend)...