2B Leveraging Data (You Already Collect) to Improve Internal Decision-Making
April 12, 2022

Caitlin McTiernan:

Welcome to day two. Welcome back. I think we can go ahead and get started right away. I know we might have some people trickling in, but getting it recorded and have a packed agenda, so we'll jump straight into things. So welcome to Leveraging Data You Already Collect to Improve Internal Decision Making. That's going to be our focus for the next 50 minutes. My name is Caitlin McTiernan, I'm a senior consultant at Just-Tech, which is a tech firm. And we specialize in helping civil legal aid groups in New York and across the country. We're really excited for the session we've got for you today. And I'm going to start off just by going through the agenda, so you know where we're headed. So in a second, I'll introduce our great panelists, and you'll be hearing from them in just a little bit. Before that, I'm going to start off with a really brief, we'll say five minute overview of the data analysis project life cycle.

We just want to make sure everybody here has a framework for the discussion that we'll be getting into. After that, we'll turn to our panelists and they'll be going through examples from their different organizations and talking about how they're using data for advocacy, for outreach, for understanding caseloads. And we'll get to those examples in a little bit. And then lastly, we'll save some time at the end for questions from the audience. And just to make sure everybody's clear, feel free to put your questions throughout the session in chat. I'll be keeping an eye at those. We are going to save those questions for the end though. So again, chat away, put in questions and thoughts and we'll address it at the very end of the session. So without further ado, I will introduce our panel.

So we have Glen Baum, he is the data and technology training manager from the Legal Aid Bureau of Buffalo. And then we have Lisa Rivera, she's the managing attorney for the New York Legal Assistance Group here in NYC. And then lastly we have Michael Mainiero, he's the project manager for Her Justice, also based in New York City. And again, we're really excited to be talking with you. We'll hear from them a lot more later today. Feel free to go camera off for the panelists for now and we'll hop back to you in a little bit. Before I jump into things, I also just want to clarify our discussion a little bit. So a lot of the examples that you hear today will be focused on data from case management systems. But certainly, the discussions we're having don't need to be limited to that. You might have data that you're collecting in other systems that can be useful for analysis. So just want to give the caveat that that's our focus, but there's definitely transferable information in this.

And then also, the focus of our conversation, what we've really worked to do is to make sure you can walk away with tangible, actionable items following this of, how to take data that you already have, maybe make adjustments to make

sure it's a better quality, it's cleaner and then have some kind of sense of where to go with it to answer strategic and programmatic questions that you might have. So I've set a bit of a timer for myself, going to try to give a brief overview of the data analysis life cycle. Obviously, sky is the limit, in terms of how long we could talk about this. It could take years of classes on data analysis. Right now, we just are going to be focused on the very basic framework, so making sure everyone on this call has the understanding of the steps to any kind of data analysis project, which you can see outlined in this slide.

So we'll jump straight into things and go through that workflow. And then we'll bring our panelists back to talk through real examples. So with any data analysis project, you'll want to start off by thinking through the kinds of strategic questions that you hope to answer. If you start out from a point of just having a data set, it's really easy to get lost. So I would strongly recommend really any robust data analysis effort is going to be driven and shaped and informed by the questions that you're hoping to answer. Sounds a bit obvious, but just a really key first step here. Depending on what your questions are, that's going to shape the type of data that you want to use, the parameters on your data, who your audience for your analysis might be and what action items might come out of your analysis. So these on the slide are all examples of questions that you'll really want to think through before you ever even start looking at data.

So for example, and this is one we'll stick with through these steps, I might be working on a legal aid organization and I might suspect that maybe we're not reaching non-English speakers at the volume that we're reaching English speakers. So that might be my question that I want to answer. Are we serving non-English speakers? Are we accessing those communities and effectively communicating our services and assisting those individuals? I think that might help me. I might suspect that I will have action items. I might have an outreach campaign that comes out of this. I might have a better sense of particular language communities to target as the results of my analysis. And I'll certainly want to include data based on our current client set of people that we're assisting. So we'll start by defining our questions.

And then next, we move to collecting the data. So most legal aid groups are already collecting a robust amount of case information, demographics, outcomes. So there's plenty of data that might be enough to just go ahead and fill in your analysis with. But it's certainly possible that you might need to adjust your data collection strategy. You might need to have built-in time to collect certain pieces of information in order to ultimately be able to conduct the analysis. So you might need to add new fields. You might need to adjust the options and fields for sticking with that language example. I might need to add options to my language look up so I can make sure that we're not grouping

everybody into another category, that I can really drill down on the types of languages that clients are primarily speaking.

Ultimately, the aim is going to be for good data quality. And by that, I mean a couple of things. So you want to have complete data. You want to make sure there's not missing values. You also want to make sure that it's accurate. So if I have a client that is a native Spanish speaker and I put English, that's just a very clear inaccurate selection. So I want to make sure I have processes in place to make sure that the data that's recorded is true to reality. And then additionally, you want to make sure that it's reliable data. So if you have multiple spots where you're collecting that somebody has limited English proficiency, you want to make sure that it's capturing the same information in both fields in a consistent manner. And also, that whoever is completing data entry has a common understanding of those fields.

An example, sticking with limited English proficiency, if you have a question that says limited English proficiency, yes, no, people might interpret that differently. So you want to make sure that there are definitions in place and that those are widely circulated so that those recording data points know how to fill that in appropriately. So we're going to define the questions, collect the data that you need. And then we get into the fun part, data cleaning. I would strongly recommend saving a copy of the raw data that you're working with, and then using that other copy to do any of the cleaning. By saving a raw file of the data as it is, whatever data set you're working off of, that just means you can go back and use it as a reference point. If you make changes to the data in the copied sheet, you can cross reference it with what it was originally.

So just make sure that you have this backup in place in case there are any issues. So you can copy your data set, have your raw one, have the data set that you're cleaning. And then you're looking for generally, to clean to make sure that it's good data quality. So address missing information, addressing incomplete information. There might be typos if there are text fields. There might be duplicative information. Maybe you want one row per case record and it's giving you multiple. So those are all steps that are involved in making sure that the data set that you're going to be working off of is clean. And then separately, you can clean to make sure that your analysis will be easier. So it's pretty frequent that people will be working in Excel when they are handling different data sets. And one thing you can do is make sure that Excel is reading the data appropriately.

So if you've got a column of dates, making sure that Excel is recognizing them as dates, similarly for number or for dollar text, so that's a piece to it. You can also remove unnecessary data, so if you've got data with maybe a hundred columns and you only really need to focus in on 25. In your defining and collecting phases, you set your sights at 20 fields. Delete out the rest of the data that you don't need, so you can just have quick access to the parameters and the set of data that you'll be analyzing. So that's step three. And as you

can clearly see, there's a lot of work that goes in before you get to this fourth step to actually dive into the data and analyze it. And your analysis is going to be driven by those questions that you set initially. There are whole classes on data analysis. Not going to be able to knock through that here, but I really just want to emphasize that summary statistics alone can get you quite far.

You don't have to be doing inferential statistics to be able to get some insights and help you answer key strategy questions that you might have. So by summary statistics we're talking about looking at trends, averages, outliers, I'm sure we all know mean, mode and median. But coming up with summaries of the data set that you have and optionally, you might choose to create some data visualizations. Those are really helpful in understanding trends. So I have bar charts, line graphs, tables alone. But again, you're not making graphs just to make graphs, you're driven by those questions that you set initially. So for example, if I'm looking into the language diversity of my clients, my data visualizations are going to be grounded in that. I might do a breakdown of primary language of clients by geographic area. I might do language breakdowns by the legal problem or by offices. But that's going to be the basis of the summary statistics and the data visualizations that I create.

And I just also wanted to flag a couple of really fantastic resources. So both New York State and New York City have open data portals. And those include easily accessible data sets on geographic areas, on a whole host of really interesting data points. And so depending on the questions you have, I could help bolster your analysis. So there are links there, and I'm sure we can send this PowerPoint around if anybody is interested. And then as the last step to this process, you're going to hopefully be able to act. So I don't want anyone to get alarmed by the word dashboards. Dashboards don't have to be a very complex thing. They could just be a collection of a couple of key data visualizations, of graphs or tables that you find the most relevant to the question.

But in however you're choosing to present the analysis, make sure that it's clear, that it's tailored to the audience that you're presenting it to, and it's self-explanatory. So if you need to put in definitions of fields within the dashboard, that could be helpful. You want somebody to be able to look at a summary of your analysis and have all the information included in that presentation to understand the data, to understand what they're looking at, to understand the parameters you might have placed on the analysis if you're looking at data only for two years. You want to communicate that so they're not thinking this is from 10 years or the entire history of your organization. So dashboards are a really helpful tool, but they don't need to be something that is incredibly complex to be insightful.

And then the life cycle doesn't really end here. So if you have action items that come out, if I've done this analysis on the language distribution of clients and I determine, oh, we need to maybe do some outreach to a particular language

group. Then I might want to collect data on that outreach and see how successful it is. I might need to add a new field in my case management system to say, "Hey, who came through? Who reached out to our organization because of this outreach activity that we engaged in?" Or I might need to collect some more data to really refine the analysis. And I might have additional questions that come out of this work.

So as I put in this last point, the cycle really starts again, it doesn't just stop there. You can go back and add additional questions, reshape your data collection strategies and keep going. I think I went a bit over five minutes, but that is the basic framework of a data analysis project. I'm going to welcome back our panelists though, so you could hear these steps in action and we'll get some examples that all of them have worked through. So I will stop my share and I will open it up with our first question. So for each of our panelists, what is one specific example of a data project that your organization has engaged in? Go in any order.

Glen Baum:

So I think I'll kick this off. Glen Baum here in Buffalo. We've been keeping data for years and years and years, all of our organizations. We have four different areas of practice and one of them being criminal defense. This is civil, but the analysis works here. So part of our criminal defense, we keep track of charges. And when the state was looking at changing the marijuana statutes, thinking about decriminalization and then ultimately, to actually get rid of the laws around marijuana or change them we thought, you know what? We've been taking data on this question for years. And of course, we keep data on our demographics as well, so we know race and ethnicities and stuff like that.

And actually, the chief of the office came to me and said, "Can you give me some data on arrests for the top charge being possession or any kind of marijuana charge as the top charge?" So I was able to go into our previous case management system, we're currently in LegalServer. But in our previous case management system we were able to go through, pull out and identify all of the cases where marijuana was the top charge. There might have been other ancillary charges, but the primary charge that got that person arrested was marijuana related. And then we were able to analyze that by race and ethnicity. 8% of all of the arrests in Buffalo were top charge marijuana related, I think in the three years we looked at, averaged around 10 to 12,000 arrests a year.

So only 8% of those were for marijuana top charge. However, of that 8%, 76% were African American. So it was obvious, this was a community that was being adversely affected by these laws. And we were able to take that and share it with the Senate and the Assembly to say, "Look, really this is harming our constituents here in Buffalo because of the way we were able to present the data and show, this is really affecting our clients." So that was one concept where we took that. And actually, part of this process also informed a bit how

we were going to keep data in our new database. We migrated to LegalServer and this process, doing stuff like this helped us to really understand how to start capturing more data. So it served a couple of purposes.

Caitlin McTiernan:

Thanks. Mike, if you want to go next.

Michael Mainiero:

Sure. I'd be happy to. So one of the things that we worked on recently was we use data, specifically geographic data, to completely revamp the way that we did outreach. We were kind of realizing that a lot of our outreach events and a lot of the way we're going about it was just, oh, what organizations do we know people at? And that's not a really good strategy when it comes to outreach. So we decided to use both the data that we had, but also the stuff that was available on NYC Open Data. So I was able to pull a couple of KML files from them and get an idea as to... And of course, this isn't perfect, but this is what we were able to pull, which is IPV incident reports via community district. We layered that along with our own data that we already collected because we keep track of clients versus city council districts, or state assembly districts, or state senate districts.

And we are able to match those and basically create heat maps and see where the really hotspots are. And then of course, we then plotted all of our current outreach events that we did in, say, the last two years. And we were able to see, all right, we've been seeing all these different places where we've been doing work. That's great. But there's this big heat map right here in the Northeast Bronx, where we have a lot of clients and we've done next to no outreach. Or we realize Southeast Queens, a very similar situation. And then we were able to tailor our approach going forward. And another example is similar to that, just using it for business processing things. Realizing that something is maybe not broken, but could be improved, and then applying data to it.

We did a similar thing with our intake program, where we realized that there was definitely issues with intakes. We solicited feedback from clients, from the attorneys that received the intake calls after the fact and the intakes themselves. And we were able to figure out what data points to be looking at, see where the problems are. Can I answer that question that just come up? Sure. Yeah. I saw a question come up and the question was the heat maps idea. So the application I used, and again, this was kind of by the seat of the pants kind of thing, was not really a planned out fully perfectly streamlined operation. But I pulled the KML files directly from NYC Open Data. It's all downloadable, it's all usable. And then I was able to add it to a Google Maps basically document on Google Drive. But again, there are much better programs for this, this was just a spur of the moment, hey, I bet I can pull this and put it

into this. But yeah, it worked out really well. I definitely recommend NYC Open Data.

Lisa Rivera:

Hey, all. So I'm Lisa Rivera from the New York Legal Assistance Group. So what I will share is, I think might be a similar sentiment for a lot of legal services organization, is that our typical relationship with data is reporting on our deliverables to funders. And we're looking at what we promised and whether we met those targets as per our contracts. This is a very admirable goal. It's literally why we exist. It's our mission to make sure that we are serving the community that we promise to. And we need to hold ourselves accountable to those that are funding us. But this relationship is pretty passive. And so our organization decided to take a bigger role in really examining what we have and what we're doing on a macro basis, but also handler by handler in each individual program. And so we did this through beginning a dashboard project. And Caitlin earlier used that word, which I didn't really know what it was beforehand, but I have educated myself.

And I think it's really helped our programs and our supervisors better manage how to design cases and understand workflow. It also will help us in the future make better projections when we are applying for new funding. What are our staff actually capable of doing? What is the normal workflow? What are the numbers and deliverables that we can actually promise? So the dashboards that we started to look at, and you could look at anything really. And so we just had to narrow the scope of what we really wanted to know. And so in the beginning, we were really looking at how many intakes are open and closed per month for each handler? What is the cross section of level of service? When we look at our client base, who's more successful in getting a higher level representation? Is there a story to be told there? Why? Is there an average for the unit to open and close in the rhythm of a month or a week, depending on their sites?

We're looking at how long on average are cases open for each type of level of service and each case type. When was it last touched? How are we defining what is active and inactive? And if it's inactive, why? Is it something that is external to us? The courts are not scheduling the court appearance? Or is our handler not reaching out to the client? Is the case just in limbo because we're waiting for a decision? We're also looking at referral sources, and I know that's something that we look at every day, but we are trying to see whether our referral sources are matching up to what community need looks like. And it's what we think, because oftentimes when we're asked, where do we think our services are going to and how, and why? And we rattle off a couple of different scenarios, but when we look at the data, it doesn't necessarily tell the same story and we try to figure out what the disconnect, if any, actually exists.

And then lastly, we're looking at what material differences between handlers in the same unit have. So in other words, is everybody getting a similarly balanced caseload and responsibilities? If no, why not? What can we learn from this? And so understanding these nuances across case handler and the program is going to better equip us for determining how and when to assign these new cases. We're going to look at numbers to figure out, what can we propose? What can we see? But also, more importantly, and I think this is something that we're all seeing right now is, the pandemic disrupted our way of work and the rhythm.

And we are often talking about before the pandemic and during the pandemic, and we're preparing for this new normal. And we don't really know what that means yet. We know what we did before, we know what we did during. And some programs really were slammed and were taking a high volume of cases. But we're functioning pretty well with that high volume. Others were much slower, but facing increasing burdens, in terms of time management, in terms of piling up cases and not being able to close them. And we're trying to figure out what that's going to really mean as the courts reopen and monitoring that in real-time. Because I think that that's going to help us prepare for that avalanche that we're all expecting to happen as the courts keep reopening.

Caitlin McTiernan:

Great. Thanks Lisa. Thanks everyone. And we'll stick with those examples that you just shared going into our next question. So what roadblocks did you encounter in your particular projects? If you want to talk a little bit too, about the steps you took to overcome those challenges, that'd be great.

Glen Baum:

So when I started analyzing the data that I was pulling to try to understand the marijuana cases in Buffalo, I noticed a concerning thing to me was that we had 16% of our cases with no race data. And I'm like, "How did we get 16% of our cases with no race data?" So I really started digging into that, trying to figure out why. And it really is for a number of reasons, nothing easily fixed because sometimes the charging documents didn't have that information and nobody thought to ask the client. So there's reasons why that happens. Another thing could be just data entry error, and that's something that we can address by creating reports to just review the data as we gather it.

So on a regular basis, running a report to make sure that all the cases have race data or all the cases have what other demographics you need to have, so that you're not getting through to the end of the cycle where you have to report on something and not having the data there. So I think that that's something that I really took away from this experience, as well as the statutes changed over time. So the Senate and the Assembly, they changed what marijuana law was.

And so now we've got a different statute, but our data entry, we didn't train them well enough on how to actually identify which statute to pick when they were putting the data into the system. So we may have had cases that were on an old statute that no longer exists and they're picking that going forward, which is a problem. So we really need to be clear. And again, this process helped me really identify ways to build in methods of avoiding similar problems in the future when we entered into a new case management system. So it really is important to analyze the data that you've collected. And this project really helped me see that in full color. Turn it over to Mike.

Michael Mainiero:

Sure. So one of the roadblocks that we came across when working with the heat maps, and working specifically with some of the information that we pulled from NYC Open Data. This is something that is very pertinent to our specific type of work at Her Justice, given the fact that we work with a lot of survivors of domestic violence. So looking at a stat like number of IPV reports per community district is not always going to be completely accurate. Because there is always the question of, people not going forward, things not being reported, things not being reported by the city itself. So we had to factor that in as well, and that's what led us to, okay, let's throw in some of our own data too, because look, we know that and we know that that's a fairly good representation.

So I guess really the point I'm making here, is really looking at things at a bunch of different angles. If you have one number, that's great. If you have five different numbers to look at that you can compare, I think that's even better. Another issue that we came across while revamping our intake system and reviewing the data from that was that one of the issues that we came across was that a lot of the complaints or a lot of some of the bigger problems were takers making mistakes during intake. While potentially putting the wrong legal problem, wrong case type or missing an address.

And that's relatively easy to track, it sounds at least until you realize that after data cleaning efforts, all those mistakes have already been fixed. And so in order to fix that, we had to pull both, not just this basic case data, but also historical case data. All right, what was the legal problem code first? What did it change to? What did it change after that? And being able to see those changes over time, let us see, okay, here are the issues that we made. Okay, here's where the actual mistakes are. Like I said, looking at just one set of data is great, but having the other set of data is fantastic. It really helps you look at a different angle and really get to what the core issue is so that you could make adjustments.

Lisa Rivera:

So a lot of roadblocks. I will start first with when we're examining such large data sets, we have to figure out whether we have the right data set to look at. And so the first thing that we had to work through, even though we have been monitoring this quite closely, is that we have a lot of cases that were inactive and that needed to be closed. And so that was our first step, sending out all of that information out, these cases look like they haven't been updated in a while. What's happening? Put a note in it if it's still alive and you're just waiting for something or whether it should be closed. Another is, I have case handlers leaving the case pending, but they're actively working on it, so it's not capturing it as an actual open case and remedying that. So we can look at it in a full data set that we can examine more closely.

And the next really embarrassment of riches, it's like figuring out what data to look at and why. And there are dozens and dozens of data points that you can really examine. And they're interesting, at least to me, a lot of them. And so I'm trying to figure out, in the world of infinite dashboards, how can we narrow the scope, so that it's useful to all of our programs? We have 14 different practice areas. We can build separate dashboards or just a couple that everybody can use because there are a lot of commonalities in the rhythm of our work is. And there'll be one or two differences, but it doesn't need to be customized for exactly every single program or unit.

And the next is a technical know-how, and so building them and learning how to keep them updated is not a skill that I necessarily had on my own, but I'm learning. And we have amazing data folks to help us get them set up and then to make sure that they're still capturing everything that we need to as we evolve and figure out how we're going to use this data to inform our practices. And then the very last thing is getting staff buy-in for these sort of processes. The last thing that supervisors want to hear is another task to their very busy days.

And supervisors have their own systems, they're very accustomed to sitting down with their supervisee, walking through an Excel sheet with a case list and hearing through somebody's cases and managing it one-on-one. What's lacking though sometimes, is that macro view of the whole team or the entire unit to see how the workflow affects every single member of the team. And so it's a different ask to say, can you put that as part of your consideration when you are managing someone? And then teaching them how to do that because it is a big difference. And then lastly, I think the other roadblock is what to do with this when we actually want to apply for new funding streams and how do we want to present this information and our findings?

Caitlin McTiernan:

So in all of your answers, I think everybody touched at least to an extent on the issue of data quality and having either missing data or maybe not collecting the data that you want in particular for that analysis. Could you talk a bit about

what types of data cleaning efforts your organization engages in and how you work to have better data quality moving forward?

Glen Baum:

Sure. So I've taken to creating reports that we can run on the data to look for obvious typos or that kind of thing. Because if I run a report looking for the intake date and the day is open and it's a negative number, obviously, there was a typo along the way there. Or if I run a report for future dates and I've got 2099, I'm pretty sure that probably most of us aren't going to be around in 2099 to be doing something. So there's obvious things that we can do in that regard, just creating reports that we can run on a regular basis, whether it's after a week of intake that you run these specific reports. And just make sure that all the data is correct and make sure that you've got all the data that you are expecting.

And then if you don't, you can flag those cases that need to be updated and send that back to the intake worker or to the handler of that particular case. So those are the kinds of things that I've been working mostly towards right now, is just trying to capture that, the need for cleanup as early in the process as possible. So when I go to run funding reports I don't have to go backtrack and start looking for data that happened six months or eight months ago. That it's already been collected. So I think that's an important thing, is to try to be proactive about that data cleaning as you collect it and not wait until the last minute.

Michael Mainiero:

Yeah. We do do some data cleaning as part of just regularly scheduled quality assurance stuff, making sure that all of our things are in order. And I've also had a couple reports set up to either go to myself or to other people within staff, showing any major errors where they're like, "Hey, here's currently where the cases are at. Does this seem right to you?" But I have found, and this might surprise some people, that not everyone likes to look at Excel files as much as I do. And as disappointing as that is, I have found that a pretty good remedy for this, a good way to just keep track where everyone is, keep track where the cases are if there's any major issues. Look, we already run so many different reports quarterly, yearly and with tons and tons of data for these grants.

And just taking a few minutes to compare that report to the previous report or look at what the numbers were previously, can give you a lot of insight. Hey, wait, all of a sudden we have a bunch of cases that were closed with no service. Why is that? And again, it's a little less proactive than, as Glen was saying, then I would like it to be, but it is something that is very helpful because it's something we're already doing. It's something that's already in front of you. And it's just a little, couple extra steps to make sure that these numbers are

within the same as where they're supposed to be from the previous year or previous quarter.

Lisa Rivera:

I'm sure it's no surprise we engage in similar tactics, in terms of a triage system so that we're flagging things on a monthly or quarterly or yearly basis. We're not, Glen, at a weekly basis, but that's something to dream towards. I will say the other thing that we do, for fields that we know that there is a high likelihood that they are not showing up or staff resistance or lack of energy. And entering certain pieces of information that are really vital for our reporting or understanding of what our workflow is. We discuss making these mandatory and concert with the teams that it's going to affect. And so we'll sit down with the program and say, "Look, we run these reports. These are the five fields, for example, that we seem to have trouble getting." And we have to do a lot of data clean up around them.

Let's talk that through. What's happening here. Why is that information not popping up? Is the field in the wrong place in your normal workflow? And should we make these mandatory? And if yes, why? And if no, talk me through what your rationale is about this. I'm really mindful that making a lot of fields mandatory is hard because you can't rush through the intake. And sometimes people are in a rush to get some information. So we do it sparingly, but we do it with purpose. And I think that we've had some success in doing that because we're talking through why we're doing it.

Caitlin McTiernan:

And that's definitely where grounding your analysis project in the questions comes into play. Like you just said Lisa, having the purpose behind why you're making something mandatory is critical, because otherwise, you risk just putting this burden on the people who are entering data to say, "Oh, we would like this. We don't know how we're using this, but we want it there so we're going to make it mandatory." You can avoid that one, you're really working to streamline what it is exactly that you need to collect.

Lisa Rivera:

And I think the conversation with why the purpose, whether it's for understanding our workflow, whether it's for a funding, we also point to the other teams that need it too. And so it's often our development team or our data analyst that needs this information. And just think, every time we're missing this and we have to go back, it slows down this team. Don't you like them too? And we bring them along in those conversations so that they understand that there's an ecosystem to the work, and it's just not them that is entering and it goes into a black hole.

Caitlin McTiernan:

So I think you've all touched on this as well, a lot of times there can be, let's say staff reluctance maybe, I don't want to say hostility, to some of this data cleaning or analysis work. So what has been your experience overall with getting staff buy-in? And then what strategies have you engaged in with particular change management to help get your data cleaned and to be able to facilitate these kinds of projects?

Glen Baum:

Sure. So as Lisa was just saying, that a lot of times it's getting the staff to understand more than just the state funders need this information. But it's how it's impacting other people in our office, other teams in our office. Or just if you didn't get that information, then somebody else has got to go back and circle around and try to dig it up. And it's usually easier upfront, so there's always a reluctance to change. And so when you start changing stuff and you start... And we did a huge change where we went from four separate case management systems into a single one, and so everybody had to change, and everybody had to learn something new. And the reluctance was huge because what you know, no matter how good or bad it is, is what you're comfortable with.

And you've learned how to work around the idiosyncrasies of that particular system. So when you go to something new, it's a matter of really almost working one-on-one in a lot of respects, to really get people to understand the importance of that data point, that one specific data point, or just all of the data points. And again, I think another thing is to listen to the feedback that you get from them because the process may become super cumbersome because you are trying to gather too much information. So taking a closer look at what is absolutely necessary and required, and then pairing it down to that. We ended up with some data gathering for one of our units that was just way over the top, way more than what, ultimately, we needed.

We took a minute, narrowed it down based on feedback from people who were reluctant, people who were really struggling with trying to use that. And so I think that's important, is to really have open communication with your staff about why you need it and then how you're going to use that in the future. How it's helping other parts of the program. And again, even with the funding, really understanding that if you want to keep your job, we've got to have accurate information to provide to funders who are giving us the money to have the job. So it's a lot of education for the most part. Turn it over to Mike.

Michael Mainiero:

Yeah. I've had a fairly similar experience, as I'm sure most people in this meeting at this panel have as well, where people do not always love to learn

more about the data or analyze the data or particularly clean the data. But what I found is, especially when it's, hey, we have this new piece of data that shows that some staff members may not be entering their cases correctly with the specific thing, which is not always the easiest thing to convey. But is one of the things that you have to do, in order to do quality assurance.

But what I found very helpful, is making sure that people are invested to a certain degree. Making sure that it's a little bit more... Make it a little more tangible for them. For example, for when we were changing up our intake things and we were looking at intake data, we were also, like I said, surveying the attorneys and the intakers that were working on it. So it could be like, "Hey, I've seen this issue so many times." And then be like, "All right, well here's what the data says." And being able to have that back and forth and turn it into a conversation, rather than just a PowerPoint, "Hey, I'm sharing this Excel file if you want to look at it." Being able to turn it into a conversation has been a lot more effective in my experience.

Lisa Rivera:

So our experience, well, time's been delightfully easy because some people get it really quickly, and other times it's like herding cats. And so I start these meetings when I'm talking about these issues, first and foremost, I acknowledge that it's burdensome and it's not something that's in the forefront of case handler minds. Because they're thinking about the case and achieving outcomes for their client on highly sensitive manners. And I think I've been successful in that because I also was a case handler and understand going through a really long intake doesn't achieve... You don't feel like that's achieving your goals. So like all the other folks, I spend a lot of time selling and explaining the value of the data. I go to each unit, I talk to the supervisors, I talk to the staff, I show them what we're learning. Trying to get them interested in the different types of data points that we're collecting because they too have an interest in learning who their clients are and what outcomes are being achieved for them.

And then one of the things that we've developed, is having good data ambassadors in each of our programs. And so those are folks that have either self-identified as really liking this, or that are pretty good at it, and we know that they are. And so we've asked them to serve as the point person for their unit, to get folks on board with entering pieces of information, onboarding new staff and volunteers, so that everybody has the same energy and the same point of view with respect to the value of this. And then lastly, the thing that not all of us talk about is, we know who our super high performers are and who has struggled putting the information in. And so I'm determined as an organization for those really high performers at the next holiday party to tell everybody that they're high performers and reward them with a nice little gift.

And for those that need it, to offer little support is like, giving them some extra training, talking to them about, this is going to be a big lift. It's a lot of data cleaning. What can we do to support you right now, whether that's extra admin support? And then let's learn for the future so that we don't have to do this again. And we've been successful in that because no one wants to see hundreds of things that they have to fix because it just adds onto their day.

Caitlin McTiernan:

I know we are coming close up onto time, so I think we can do a really rapid fire last question. For our panelists, what would be your recommendation? Your one recommendation to the audience for how they can start leveraging their data better today?

Glen Baum:

So I think a couple of things, you really need to take stock of the data that you're collecting and analyze, why are we collecting it? So really, that's I think, the biggest, most important thing for me, is to really understand what data you're collecting and why you're collecting it. And then from there, you can start figuring out how you're going to move with it. But that's the most important thing, is to know what you're collecting and why.

Michael Mainiero:

Sure. I would say just in a very, very general sense, be proactive. Be proactive in your data cleaning, be proactive in, hey, this particular process isn't working the way we want it to work. Or hey, this could be improved. I guarantee you that there is some data in there that can tell you exactly why it's not working or exactly how it can be improved. Be willing to look at other sources like NYC Open Data. Be more proactive and less reactive.

Lisa Rivera:

I think knowing what you have is really important, Glen, I'm so glad that you said that. And I think the other thing is re-examine your values. Why are you serving X population? Who do you want to serve? And does it match up? And if it doesn't, why? What are the reasons? And try to align what you think you're doing, to what is actually happening. And you can do that very richly and easily through data and then slowly make benchmarks for yourself to get to your ultimate goal.

Caitlin McTiernan:

Fantastic and fast. Thank you. I'm going to share my screen again just to go over a couple of other takeaway items. There's a lot of text in there. You can grab a screenshot of it. And I'm sure this PowerPoint will be accessible if you want it. At this point, I know some people have put questions throughout. If

anybody has any questions at this time that they want to add, feel free. We've got a couple of minutes left, so happy to answer. And I do see one from William. So this is particularly for Glen, in the process of using your data points in regards to cannabis arrest, did you work with external partners through the whole process? Or did you see the findings and then find the relevant external partners? So we'll start there, there's two parts.

Glen Baum:

So to be honest with you, I was approached by the chief attorney for the public defender program and asked for this information. I gathered the information and handed it over to him. And so I don't really know what partners he worked with. I wasn't part of that part of the policy experience or process. I basically pulled data out of our system and handed it off. So I wish Kevin were around then I could talk to him about that, but I can't really answer your question there unfortunately.

Caitlin McTiernan:

I think that's an example, that was clearly translated to publicly distributed data set. I don't know if anybody else, Lisa, or if Mike, if you have any thoughts on managing the policy side of data initiatives or working with external partners and maybe publicizing data externally. I don't know if you have experience or if you have any comments on that example.

Lisa Rivera:

I think similarly, as Michael was talking about, using the heat maps to figure out where your services are, and to determine where need is. And so we've been able to point out to where our high COP qual volume is to certain areas of law. And then use that in our advocacy, whether that is either pitching that to a funder for an idea, or going to city council and explaining why we're asking for increases in initiatives or a brand new initiative to point to need. So at a bill that's being passed and we try to use that data richly and support it with that larger New York City, local information.

Michael Mainiero:

We do similar work as well. The elected officials that work with us they are also very invested in our work. They care because their constituents are the ones who need these services. So we wanted to also keep track of not only, all right, where are these heat maps? But also, where are these city council districts? Where are these state assembly districts? Because these people clearly have an investment in this. We need to be on top of this as well, and make sure that we're tracking that. And also, make sure that we're doing outreach in those areas because clearly there's a need.

Caitlin McTiernan:

And Sally, this goes back to your question about the applications that you asked at the beginning. So I know Mike, you highlighted the open data portals for both the city and the state have political geographic data that you can use for those that are LegalServer users. LegalServer has modules, I believe they come with an extra cost, but they can geocode. They can map your data based on the address information, which gets you a bit of a head start.

But for those that just have addresses, maybe they don't have the coordinates in any system. There are publicly available tools, I think even the Census supports one, that can geocode your information. Might take some cleanup, might be a little bit slower of a process. But no matter what format you're tracking geographic data in, there are publicly available and free tools to map that data. And then I just also wanted to highlight QGIS is a platform that's open-source for mapping. So you can definitely take the open data from NYC, you can take your geographic data from a case management system, combine those all together in a platform like QGIS and get your heat maps. There might be easier solutions too, I think Excel has some as well.

Michael Mainiero:

I was going to say, I'm sure that that is a much better system than what I put together on Google Maps. But yeah, definitely use those things that are available to you, make sure that you use those different things that you have.

Caitlin McTiernan:

And a great point, if you can get Google Sheets, if you can get Excel to do some of this work, complexity of the system doesn't have to mean that it's better. You can get insights with less robust tools, we'll say. Yeah. I know we're going a bit over, any last questions? Do I have a website for the GIS... Sure. Yeah. Let me go ahead and grab that one. I'll put it in chat and then I'll do a last call for any questions from the audience before we wrap for today.

So this is the link to QGIS and you can download it. And then I'm actually not sure I have a geocoding off the top of my head, but if you Google it I'm sure there are multiple platforms that support it. Okay. Well, a huge thank you to our panelists. That was a fantastic discussion, and I think incredibly insightful. Hopefully, everybody can walk away with some ideas of where to go next with their data. And thank you to the audience too for sticking around with us. We hope you have a great rest of the conference. I'll also throw my email in the chat in case anybody has any questions and wants to reach out, specifically on any of the things we talked to today. So thank you, everyone. We'll see you throughout the comments.

Glen Baum:

Thank you all.

Lisa Rivera:
Have a good day.

Michael Mainiero:
Thank you.