# Getting Started with Secondary Data Analysis with Kate Guastaferro and Loren Masters

June 11, 2017

Aaron Wagner: Methodology Minutes is brought to you by the Methodology Center at Penn State. Hello and welcome to Methodology Minutes. I'm Aaron Wagner, your host. With me today are Dr. Kate Guastaferro and Loren Masters.

Kate is a postdoctoral fellow in the Methodology Center, Prevention and Methodology Training or PAMT program. In PAMT, she is working with Jennie Noll as her prevention mentor and Linda Collins as her methodology mentor. Kate got her PhD in Public Health from Georgia State University. Her research focuses on the development of behavioral interventions and child maltreatment.

Loren Masters is a data manager and analyst at the Methodology Center. Loren has a master's degree in Public Health, with a concentration in epidemiology and biostatistics from Penn State Hershey and is responsible for managing complex data sets, initiation, and maintenance of data set and compliance documentation and performing statistical analyses in support of the center's investigators' research at the Methodology Center.

They're with me today to discuss the process of establishing a secondary data analysis research program. Coming out of graduate school, many researchers have worked exclusively in their advisor's labs. Many young researchers look to secondary data as part of their research portfolio, which is great. Secondary data analysis is a high priority for many funding agencies, as they try to get the most out of studies they have already funded.

Almost everything we do here at the Methodology Center is secondary data analysis, but there are barriers to getting started. While this podcast cannot remove those barriers, it can help you think about what they are and how to address them, so thank you both for being with me today.

Kate G.: Thanks for having us.

Loren Masters: Thanks, Aaron.

Aaron Wagner: Kate, prior to joining the Methodology Center at the beginning of this school year, did you have any previous experience with secondary data analysis?

Kate G.: Not really. My experience is in intervention development and evaluation, so I focus primarily on primary data collection and analysis that way.

Aaron Wagner: I understand that now you are working on several analyses on data sets from other universities, is that right?

Kate G.: Oh, yes.

Aaron Wagner: A lot of researchers work with what we call restricted data that is available for qualified investigators. Can you talk about what went in to getting permission to work with data from one of these, the National Survey on Child and Adolescent Well-Being?

Kate G.: Sure. Well, first of all, the data set is a nationally representative, longitudinal study on children and families who have been involved in Child Protective Services. It's managed by an office at Cornell.

The first thing that I did was see if anyone else wanted to work with the data, my colleagues in the center, and I found another postdoc who was interested in working with it. We met with Loren and determined the steps that we needed to do to acquire the data and after that, we worked with her to create an IRB application to work with ... It was a protocol for analysis of a restricted data set, and then after the IRB was approved, we did another application because that's how this goes, right?

The application was from Cornell. I had to provide an analytic plan, a scope of work, and the rationale for using some of those variables, and then how we were going to manage and protect it with Loren's part.

Aaron Wagner: Just what you've described there, it sounds like a huge amount of work. How long did that process take?

Kate G.: A couple of months maybe to get the IRB straight and make sure that we both had an idea ... My colleague and I had clear ideas of what he would do versus what I would do.

Aaron Wagner: Okay. My guess would be that the IRB process differs pretty substantially from university to university? You're shaking your head, which we can't hear. It's pretty standard, you're saying, from university to university?

Loren Masters: Yeah. The standards for how you conduct an IRB, usually, the accreditation is received nationally and the criteria for certain protocols and how you manage an IRB and what approvals are necessary, it's actually mandated by the federal government, so it's pretty straightforward across universities.

Aaron Wagner: As you move from place to place, you can expect some consistency in the process. What about the timeline?

Loren Masters: I'd say that most large universities have a probably similar timeline and if you are requiring an IRB approval, that requires that the IRB board meets. When they reconvene each month, they'll go over the protocols that are on the agenda, so I say that establishing that relationship with your IRB analyst is probably one of the first steps that you need to establish at your new university. IRB analyst is someone who will work with you to navigate the process of an IRB.

In Kate's instance, for this particular data set, she did not need a full IRB approval. It was an expedited approval, which means that the IRB analyst was able to review it with a few other people with her expertise, and they were able to give Kate approval without having to go to the full board.

Aaron Wagner: For someone who just wants to do research, maybe this sounds a little complicated, frightening and frankly, awful. Is it?

Kate G.: No. Like Loren said, it's a standard process, so you kind of know what to expect from A to B. I would say, what's different about Penn State from my former institution is the procedures are very well-laid-out here, so it wasn't particularly scary in any way. Plus, I'm extremely lucky and have Loren as my data manager to help me walk through those steps.

Aaron Wagner: If somebody is saying, "Data manager? There's no data manager here," does that mean it's incumbent upon them, the investigator, to get this done on their own?

Loren Masters: Yes. At some other universities, where there isn't someone like a data manager ... In our center, it's a little bit different because we have the infrastructural support that helps alleviate that research burden, but in other places and other universities and even other departments at Penn State, they don't have someone like a data manager, but there are people like those IRB analysts that are, I would say, probably consistent across most units and they are always there to help you regardless if you have a data manager or not.

Aaron Wagner: That's terrific. Okay. I interrupted the flow of that. We talked about meeting with the data manager and to set the steps. You talked about the IRB, and then there was an application for the restricted data set, right?

Loren Masters: Yes.

Aaron Wagner: What comes after that?

Loren Masters: Okay. The next step, after the application for restricted data, is the data protection plan or a data security plan. There's a few different words for it, but basically, this plan is different from the application for restricted data in a sense that it talks about how you secure and protect the data, agreeing to certain conditions and that you'll do so.

How you do that is you have to work with your IT staff. Usually, at Penn State, there are centralized staff and there are also local staff related to your college. For us, there's an Office of Information Security and there's probably some other compliance office that's equivalent at other universities, and those staff work with your local IT staff within your college or department, and they will help you determine certain components of data security that you will need to adhere to while protecting the data.

Usually, that involves about four components. Physical security such as a locked office that's private. Electronic security, something like your password and having a complex password. Access control, making sure that only investigators who are allowed to have access to the data have access to it, and administrative security, making sure that certain administrative controls are placed on the data with regards to the contract.

Aaron Wagner: Anyone who's worked with data in any form is probably familiar, at least, somewhat with these sorts of things.

Kate G.: Yes, but can I add that the access control and physical security is, as a postdoc, something I couldn't control on my own, so I needed to have a specific office that locked with the door, that my computer with these data could be safely housed on.

As a postdoc, I don't have an office like that, so this required more in helping me get the buy-in and the permission from leadership in our center to have a physical space, so it really is a cooperative process. You need IT, but you need your mentors and a lot of folks.

Aaron Wagner: It takes a village.

Kate G.: It does.

Aaron Wagner: Okay. Once we develop the security necessary to have the data ...

Loren Masters: Right. After you've developed the protection plan, it's approved by, usually, your institution first, and then once it's approved by your institution, it goes to the data provider and they also have to approve that plan. You can make some modifications, but that's why it's important to have your IT staff involved throughout the process.

Once the data protection plan is approved, your restricted data license that is also part of the package that you have to submit and get approved to the data provider that requires a legal signature from your institution that you're at, and that's something that's unknown to newer investigators. You may not know that you're not allowed to sign these legal documents until you are a part of a big process like this.

What that means to a data manager or someone like Kate is that the document, which is a data use agreement or a restricted data license, something like that, has to go to a variety of offices at Penn State or your other institution, and they will review the agreement and make sure that the terms and obligations that are a part of the contract can be fulfilled.

Once they do that, they have certain officials who are allowed to legally sign the document, and there are three offices that do this at Penn State. Some institutions may only have one office that does this, but Penn State has three, and it depends on the cost and the fee for the data, and it also depends if there is grant funding the data acquisition or if the research is exploratory.

Typically, the PI or a post-doctorate graduate student is referred to as a data steward. In Penn State and other institutions, we'll still consider them to be obliged to the terms and conditions that are a part of the agreement, but if there were something to happen, like a breach of the agreement, Penn State usually assumes a lot of the liability with those terms and conditions.

Aaron Wagner: All right. Thanks a lot. That's kind of a lot to wrap your head around.

Kate G.: It's intense.

Aaron Wagner: Yeah. Other than that type of data, are you working with any restricted but available data sets other than that one?

Kate G.: Yes. I'm very lucky to be at the Methodology Center, where we have several other available data sets, and Loren helped me get added to several of those agreements to use these data sets that we already have permission for. Loren, do you want to talk about that, since you worked all behind the scenes for me?

Loren Masters: Sure. If there's already a data use agreement that's established with another investigator, sometimes, it's easier to add junior investigators or other investigators who would like to be collaborators on certain projects. As long as you follow within the same scope that was proposed and the data use proposal, then usually, it's pretty easy to add additional research staff.

You also will need to make sure that you have an IRB that also falls within the scope of the research that was originally proposed and most likely, you'll need to sign a form that's a research staff agreement of some sort and agree to certain terms.

Aaron Wagner: Okay. As we talk about these different categories of data, what distinguishes a restricted data that is available for qualified investigators and how do researchers identify what might be out there that they could access to do secondary data analysis?

Loren Masters: There's a lot of different types of data, but I'm only going to talk about two today to try and simplify the process. What Kate was trying to acquire was what we call a restricted data set that's available for qualified investigators, and that data set is different than a public data set in a sense that someone that's in the general public can't just go click on a link and download the data.

With the restricted data set, this is a data set that usually requires some type of approval, an IRB approval. It usually includes a data license or a data use agreement, and it requires legally signing and agreeing to terms and conditions.

Aaron Wagner: Okay. Thanks. Kate, like a lot of people who are launching a independent research career, you brought some projects from your prior institution. Was the process of obtaining that data easier, harder or much the same than the process we've just described?

Kate G.: I would say it's easier on the whole but entirely different.

Aaron Wagner: Okay.

Kate G.: Of course, it requires IRB approval, but because these projects were ongoing data collection that I had that involved with my prior institution but wanted to be able to use the data here, the data I brought were active projects at my prior institution, so I had to go through an IRB process here, which, like Loren said, has similar elements across any institution, but because they were active at a prior place, I had a different kind of application, in which I can never remember the name of.

Loren Masters: Non-human subjects research.

Kate G.: Oh. Is that simple?

Loren Masters: Or not engaged?

Kate G.: That's the one. Not engaged. I was not involved with the data collection or recruiting participants, which meant it was a different kind of IRB than I had been accustomed to, previously.

It also has to do with the stage of the research, so some of the projects had ended and I just wanted to be able to use the data to publish. Others were still active and I had a role in providing some consultation, but I would say it's easy. However, it's easy because I have Loren.

My first week here, we sat down in her office and I brain dumped everything that I was bringing with me, and she asked very thoughtful questions about what kind of data, what I was going to do with it, and then helped me have a strategic plan on how to get permission to use that data.

Aaron Wagner: Okay. Well, as we alluded to you earlier, most people don't work with Loren.

Kate G.: I know. How unfortunate.

Aaron Wagner: Yes, it is unfortunate for them. Basically, we're just doing this podcast to rub people's face in it, but if someone didn't work with Loren, what should they do? They should brainstorm on their own, presumably everything they want to work with, right, and then what?

Kate G.: My suggestion would be to talk to an IRB analyst. We have people trained to do this and they know the right steps, and so you don't have to figure it all on your own. There are resources even if it's not Loren.

Aaron Wagner: Great. We wanted to end with a cautionary tale. Can you explain your data procurement horror story, the one project that has given both of you fits ever since you arrived? We're recording this in April, so...

Kate G.: I got here in June last year.

Aaron Wagner: Yeah, it's been about a year.

Kate G.: It has. It's a horror story, but I think cautionary is correct. What this entails, as I was trying to bring a data set that I had used for my dissertation and had permission to work with my graduate program but needed to bring it here, the unfortunate part about this data set is that it's a little bit older and the PI who had collected it had passed away.

I had a lot of support and buy-in from people who wanted to publish off this data and were excited about my ideas, but it's not their project, and so it was harder to get permissions and this involved, as we just said, from June to April, lots of legal processes to get agreements between the two schools, where the data was housed and collected versus where I am.

My word of caution is do not put all of your eggs in one basket and hope that you're going to get that data quickly because you don't know what's going to happen, and there are lots and lots of things outside of your control, and so just diversify.

Aaron Wagner: Excellent. Thank you. Research advice is like financial advice.

Kate G.: Yeah.

Aaron Wagner: In the end, what should people take away from this podcast? What should researchers do and when should they do it?

Loren Masters: I've been able to have the opportunity to meet with some of the other trainees in the PAMT program that Kate's in, and I always tell them the same thing. The earlier that you can anticipate needing or acquiring data, the better. Usually, the process does vary for every institution, depending on what resources you have, what leverage you have for getting an agreement processed quicker, so a lot of preparation is needed on the front end.

If you're on top of the preparation on the front end as an investigator, then you can make the process go much faster because usually, once you finished all of the IRB, the data use proposal, all the things in the front end that Kate and I talked about, after that part, usually, what's needed is only the negotiation and the signature, and that could take months at least. If you can finish that front part, if you could condense in from six months to two months, that'll save you four months there.

Also, make sure you allow a lot of wiggle room for the offices at whatever university you're at. Each administrative office may have a lag time of four weeks to six weeks to six months. Make sure that you're aware, whenever you're not using primary data and it is secondary data, if you are downloading data or if you notice that there are terms that you're about to agree to, just let that be a red flag and remind you that there are people at your institution who are likely more qualified to help you with that than you're just trying to handle it yourself.

Aaron Wagner: All right. You got this one on podcast aggregator or something like that. Go to the Methodology Center website and find the cheat sheet as well because if you're interested enough to listen to a podcast, you should definitely look at Loren's notes. Well, Kate and Loren, thanks very much.

Kate G.: Thanks for having us.

Loren Masters: Thank you.

Aaron Wagner: You have been listening to Methodology Minutes. Please visit methodology.psu.edu for podcasts, software, reading lists, and more.