

Time transcripts of Panel_2_Roundtable

[00:00:00:020] >>MOLLY IRWIN: Um... Before we do, I was just reminded

[00:00:04:030] by Naomi that today is OPRE

[00:00:08:040] Erica Zelewski's birthday, [laughter] so everyone

[00:00:12:060] should join me in wishing Erica a very happy birthday.

[00:00:16:070] [laughter] [applause}

[00:00:20:110] [inaudible over applause] ...and Erica is right back there. Um...

[00:00:24:130] I'm Molly Irwin from the Chief Evaluation Office in the Department of Labor

[00:00:28:150] and I'm really happy to be here this morning and to be moderating this

[00:00:32:180] roundtable discussion on gaining access and roundtable discussion on gaining access and maintaining

[00:00:36:210] confidentiality with administrative data. I think, um,

[00:00:40:220] this session will be a really nice follow up

[00:00:44:220] to the conversations that we've already had this morning, and I think will be

[00:00:48:240] a nice way to talk about some of the

[00:00:52:260] the... um... sometimes challenging issues

[00:00:56:290] and hopefully really good solutions to gaining access to data,

[00:01:01:010] so that we can move on and use the data

[00:01:05:020] in the cool kind of applications that I think we're going to hear about for the next

[00:01:09:030] day and a half, so I think this will be a good stage setting

[00:01:13:040] session. It's going to be a roundtable discussion

[00:01:17:060] we have four great panelists

[00:01:21:100] and a discussant. I'll introduce of them and then we'll go right

[00:01:25:110] into it and then have some time at the end for discussion with the audience

[00:01:29:130] So... the first person on the panel is

[00:01:33:150] Beth Green, who's the Director of Early Childhood and

[00:01:37:180] Family Support Research at the Center for Improvement

[00:01:41:190] of Child and Family Services at Portland State University. Her work focuses

[00:01:45:190] on managing, designing, and implementing studies of early childhood

[00:01:49:200] education and parenting programs, children's mental health,

[00:01:53:230] early literacy, child abuse prevention, community development

[00:01:57:270] coalitions, child welfare, and family drug courts.

[00:02:01:280] Next we have Chuck Michalopoulos who's the

[00:02:05:290] Chief Economist at MDRC. His recent work includes

[00:02:10:000] serving as a co-PI on two national evaluations of home-visitng

[00:02:14:020] programs, as well as leading to evaluations of coordinated

[00:02:18:040] care for high-cost Medicaid recipients. All of these studies

[00:02:22:060] involve accessing and working with administrative data from different state agencies

[00:02:26:070] across multiple states. Maya Bernstein, right

[00:02:30:080] to my left, is a Senior Policy Analyst and Privacy Advocate

[00:02:34:100] at the Office of the Assistant Secretary for Planning and Evaluation

[00:02:38:150] in the Department of Health and Human Services. There she advises on

[00:02:42:190] legal and policy issues involving the use and disclosure of personally

[00:02:46:200] identifiable information,including how and when it's appropriate

[00:02:50:210] to give researchers access to administrative data.

[00:02:54:240] Jennifer Noyes is the

[00:02:58:250] Associate Director of Programs and Management at the Institute

[00:03:02:280] of Research on Poverty and the Co-Director of the Center on Child

[00:03:06:290] Welfare Policy and Practice at the University of Wisconsin, Madison

[00:03:11:010] Her work involves using administrative data within the state of Wisconsin, as well as

[00:03:15:030] across multiple agencies and states, including efforts to create

[00:03:19:060] files that link across administrative data sets. And finally,

[00:03:23:070] Kelly Maxwell, who will be our discussant, is a

[00:03:27:080] Senior Research Scientist and Co-Director for Early Childhood

[00:03:31:090] Development at Child Trends. She's known nationally for

[00:03:35:110] her policy-relevant research and evaluation focused on helping states

[00:03:39:150] improve their early care and education systems.

[00:03:43:190] So, I think we are well set up to have a good discussion.

[00:03:47:200] What I'd like to do to start out is really just to go down

[00:03:51:210] the row and ask all of the panelists

[00:03:55:230] just to give a couple sentences, a high-level overview

[00:03:59:270] that describes their relevant experience working with

[00:04:03:280] administrative data. [inaudible]

[00:04:08:150] We'll start this way. >>CHARLES MICHALOPOLOUS: Um.. I mean, a couple of sentences, it's...

[00:04:12:220] So that, uh, yeah... A couple of studies were mentioned; two that we're doing right now are

[00:04:16:230] evaluations of home visiting programs where we're collecting

[00:04:20:260] three types of administrative data from seventeen different states plus data from the

[00:04:24:280] new hires database. So that'll be the basis of a lot of what I'm saying, but I've been working with administrative data

[00:04:29:020] for a couple of decades, so I'll bring in

[00:04:33:020] some of those experiences as well.

[00:04:37:040] I think we'll be reinforcing a lot of what you heard in the previous— y'know, the previous panel.

[00:04:43:290] >>JENNIFER NOYES: So, as is mentioned in my introduction, I spent a lot of time

[00:04:44:010] >>JENNIFER NOYES: So, as is mentioned in my introduction, I spent a lot of time within our state working to integrate data

[00:04:48:030] across multiple state agencies and have been leading the development of what we call

[00:04:52:070] the Multisample Person File, which integrates data across about twelve state agencies, including

[00:04:56:070] the entire populations of.. of.. um.. included in that data.

[00:05:00:080] One of my main functions is I work with other researchers to coordinate their data requests,

[00:05:04:100] and to help get access, and get all the confidentiality agreements

[00:05:08:130] and those types of things in place. I'm also involved in a similar

[00:05:12:160] cross-site national study that duplicates some of Charles's experience

[00:05:16:170] but most of what I'm going to talk about today will draw on my in-state experience working across, as I said,

[00:05:20:180] multiple state agencies, ranging from Department of Corrections, Department Workforce Development

[00:05:24:200] Department of Health Services, Department of Children and Families, our Department of Revenue

[00:05:28:220] with IRS records; that type of a thing. [microphone bump]

[00:05:32:250] >>MAYA BERNSTEIN: Hi. So, I'm a— I think have a little bit different

[00:05:36:250] role than the other people on the panel. I'm not a user of data;

[00:05:40:260] I'm a gatekeeper, I guess. [laughter] And so my perspective is going to be

[00:05:44:280] a little bit different. I guess that's why I'm here Uh...

[00:05:48:300] I spend a lot of time working with people who are trying to get access to data.

[00:05:53:010] Um... and I, uh, In spite of my title, I tend to be very pragmatic.

[00:05:57:040] But I want to make sure that the data that people

[00:06:01:050] are seeking—that there's a nexus between what they're trying to seek and

[00:06:05:080] what they're actually trying to accomplish. And, often times we find

[00:06:09:100] that people are trying to access data that is not the right data for their

[00:06:13:130] problem that they're trying to solve or the research that they're trying to do.

[00:06:13:160] problem that they're trying to solve or the research that they're trying to do. Um, if I...

[00:06:17:170] And I... You mentioned the National Directory of New Hires; I started my

[00:06:21:180] federal career at the Office of Management of Budget, where we cleared the legislation

[00:06:25:200] and the paperwork— all the things that set up the National Directory of New Hires,

[00:06:29:210] and since then have been reviewing all of the many requests for access for it

[00:06:33:240] And so, I'm always looking for the nexus between what the data is

[00:06:37:260] that we have, and what the research problem is, or research question is

[00:06:41:270] or the administrative problem that someone is trying to solve, and making sure that that's rights.

[00:06:45:280] If that's there, I'll go to bat for people, but if it's not there, you know

[00:06:50:020] there's not a very good argument for access to the data.

[00:06:54:050] >>BETH GREEN: And, um... I have had, most recently, a couple experiences with administrative data.

[00:06:58:080] I'm primarily a program evaluator and researcher. So, take coming at it

[00:07:02:090] from that side— one of the people trying to get access. Um...

[00:07:06:110] Using it primarily for program evaluation experience in a variety of different contexts.

[00:07:10:140] The two most recent are a

[00:07:14:180] retrospective evaluation of the early Head Start national

[00:07:18:200] national study, which was done, um— it started in the late nineties,

[00:07:22:210] actually, and involved three thousand children across seventeen states.

[00:07:26:240] And about seven years ago, we received some funding through

[00:07:30:270] some various state agencies to go back and see if we could dig up

[00:07:34:280] Child Welfare administrative records for those three thousand families.

[00:07:38:290] Um... and go retrospectively— So using administrative data

[00:07:43:010] that way. And so we've gotten now seven—we've got data originally

[00:07:47:030] in the first study from seven different states, and we're in the process of accessing that data

[00:07:51:060] from the remaining ten states currently.

[00:07:55:080] The second study involves a randomized trial within the State of Oregon, where we're looking

[00:07:59:090] at the effects of a home visiting program and linking data through seven different administrative

[00:08:03:100] data sets within the state. So, very different experiences.

[00:08:07:130] >>MOLLY IRWIN: Great. With that let's start now to really

[00:08:11:140] um... get each of your views

[00:08:15:150] about what the biggest challenge you faced

[00:08:19:160] in accessing administrative data, understanding

[00:08:23:180] what it actually means, or maintaining confidentiality.

[00:08:27:200] And those are all really big questions that we'll dive into a little bit deeper,

[00:08:31:230] but maybe just a high level overview from each of you.

[00:08:35:250] And maybe we'll start with Beth this time and then head that way.

[00:08:39:260] >>BETH GREEN: So... It's a little bit difficult to say what the biggest challenges are.

[00:08:43:280] I felt like every single challenge we've had was mentioned by

[00:08:47:300] Dr. Patrilla this morning when he was going through his list. I was like, "Oh yeah, that's happened. Oh yeah that's happened too."

[00:08:52:030] "Oh yeah, that's happened too." Everything from people saying

[00:08:56:040] we can't do it, because it's against our legal statutes;

[00:09:00:040] we don't have the resources to give you the data you want;

[00:09:04:050] to challenges once you get the data in storage and security and analysis.

[00:09:08:070] I would say... [sigh] I would say the—

[00:09:12:100] If I had to pick the biggest challenge, it would definitely have to do

[00:09:16:140] with working with the state agency partners,

[00:09:20:150] who vary tremendously

[00:09:24:170] in their capacity to respond to these kinds of requests,

[00:09:28:190] as well as their level and sophistication in doing so. We noticed huge, huge

[00:09:32:200] shifts, even from the first phase of our early Head Start study, which we did

[00:09:36:250] about seven years ago, in the level of sophistication,

[00:09:40:260] which is good, around state agencies thinking about how to

[00:09:44:270] create data use agreements and data sharing agreements

[00:09:48:300] This is specific to child welfare administrative data. But, the resources and

[00:09:53:030] the capacity at the state level has been a huge challenge,

[00:09:57:030] and even trying to deal with that challenge, by building money into our grants

[00:10:01:040] to provide funding to them to release staff time to help us respond

[00:10:05:060] to the requests— things like that. In many cases, it's just

[00:10:09:080] a matter of time and trying to get to the top of their

[00:10:13:120] queue, in terms of responding to your request. So, that's probably been the

[00:10:17:140] biggest challenge, I would say.

[00:10:22:120] >>MAYA BERNSTEIN: Um... [microphone thump]

[00:10:26:140] The biggest challenge for me? [laughter] Uh...

[00:10:30:150] I think, um...

[00:10:34:180] Well, one of the things is what I mentioned before, which may be a theme throughout,

[00:10:38:220] which is making sure that the person— that we know

[00:10:42:220] what the data is and what the problem is; and, defining that problem

[00:10:46:240] and defining the nexus between the data as I mentioned. But, I think, um...

[00:10:50:250] I try to think about...

[00:10:54:280] when people are trying to access administrative data,

[00:10:59:010] the kinds of studies that heard about, the kinds of things that you're doing.

[00:11:03:020] Many of these administrative records are about particularly vulnerable populations.

[00:11:07:040] People have to give us a lot of information in order to get the services that

[00:11:11:050] we're providing. And for me, I like to try to remember

[00:11:15:090] and to try to remind the people that I'm working with that this is—

[00:11:19:090] That the... that the use of data for various kinds of research purposes

[00:11:23:100] or other administrative purposes that are different from that from which they were

[00:11:27:120] originally collected is potentially another vulnerability

[00:11:31:140] for this population. If we don't carefully safeguard the data; if we're not

[00:11:35:170] concerned about how we're using the data and how it's reported,

[00:11:39:200] just another vulnerability. And so, I try to

[00:11:43:210] remind people that underlying all the work that they're doing— yes we're doing good and

[00:11:47:230] yes we want to improve outcomes, and all the good things that we're doing, but

[00:11:51:250] we don't— we're not entitled to this data. Most of the data is given to us

[00:11:55:280] voluntarily by people who want to

[00:11:59:280] get services or... y'know... if you think the position

[00:12:03:290] people are in, maybe not so voluntarily. And, so I like to try to

[00:12:08:010] remember to be respectful of that when we're thinking about using the data;

[00:12:12:030] and, to be careful, and to try to use the minimum amount we need

[00:12:16:050] to get the job done; and to be— and, and

[00:12:20:090] to remind ourselves that if we— if people are not confident

[00:12:24:100] about what we're going to do with the data and how we're going to use it and how we're going to

[00:12:28:120] report our results; they might not give us the data, in the first place, and then we're nowhere.

[00:12:33:240] >>JENNIFER NOYES: So, my challenge builds off of that

[00:12:37:250] but in a bit different direction, and it is— the challenge has been mentioned

[00:12:41:260] several times already this morning— that the data that we're trying to access and use was not created research

[00:12:45:280] purposes, and it was created for other purposes. And one of the challenges that I have

[00:12:50:010] faced in working with and trying to be the bridge between some of the researchers

[00:12:54:040] with whom I work in the state agencies, is to get the researchers to understand that

[00:12:58:040] it's not a one-way street. That going into ask for data that

[00:13:02:060] they may not understand, or that wasn't developed for the purposes they want to use it for,

[00:13:06:090] and they need to invest time to understand it, requires a reciprocal relationship

[00:13:10:120] back to the agency owners or the agency providers; that there has to be something

[00:13:14:160] in it for them. That it can't just be

[00:13:18:160] "please let me have your data, and I'll walk away and never speak to you again about it."

[00:13:22:180] Because, otherwise you can't get the buy-in, In my opinion— the long-term,

[00:13:26:200] over-time buy-in. I call the two essential ingredients

[00:13:30:220] the two T's: the trust and the time. That if you aren't—

[00:13:34:250] There's an agreement that you have in terms of confidentiality and privacy

[00:13:38:260] with the human subjects that provided the information, but then there's the agreement

[00:13:42:290] I believe that you have to have with the agency owners— the data owners

[00:13:47:010] to give something back to them for what it's going to take for them to give it to you.

[00:13:51:070] So if it's a one-way street, it's not going to work, and it's certainly not going to work over time.

[00:13:55:080] An if you want to develop long-term relationships, you have to spend that time to develop that trust.

[00:13:59:090] If it's just a one-off, you want to pull the data once and walk away I don't think that's healthy

[00:14:03:120] or smart, and that has been one of the biggest challenges, I think,

[00:14:07:150] in terms of working with researchers who want to use the data, and then also convincing the agencies to

[00:14:11:180] get in the game with us. >>CHARLES MICHALOPOULOS: Um... I, uh, I...

[00:14:15:190] I agree with all the challenges that have been talked about, and I think we face them as well.

[00:14:19:210] But if I had to— yeah, since we were asked to pick a biggest challenge,

[00:14:23:230] I would say, especially in the context of a federal study, like the home visiting studies

[00:14:27:240] we're doing, time is the biggest challenge— the time it takes to accomplish all the things

[00:14:31:280] that are required. I mentioned that we're working with two or three

[00:14:35:280] agencies to collect Medicaid data, birth certificate data,

[00:14:40:000] and child welfare data across twelve states for Child Welfare

[00:14:44:040] and seventeen for the other two data sources. So it's quite a few state agencies.

[00:14:48:080] And, yeah, as we've heard, there's a multi-step process of contacting the agency,

[00:14:52:090] developing a relationship with them so they understand what you're, you know, what you're trying to

[00:14:56:100] get, and that they're interested in helping you out. But that's just the

[00:15:00:130] easy part, in some ways. Because, then there's a data committee or an IRB that has to review an

[00:15:04:160] application. Um... They'll have to review a consent form that you've spent

[00:15:08:160] a lot of time carefully crafting for your national study. And they'll

[00:15:12:160] probably want changes to. Um... And...

[00:15:16:170] Then finally there's a data use agreement which, you know,

[00:15:20:190] has to be negotiated. In part, because, again

[00:15:24:210] for these two federal studies there are restrictions on what we can do with the data from the

[00:15:28:250] federal government's perspective. Those don't always agree with how the states want to use

[00:15:32:250] to use the data. So, negotiating those has definitely been time consuming. I think over all,

[00:15:36:260] I think, over all, um... we've signed agreements with most of the agencies

[00:15:40:270] we've talked to, but it's taken, on average, a year– a year and a half.

[00:15:44:290] probably for each agreement to get signed. So, within the context of

[00:15:49:040] a four or five year study, which is more the ones we're doing, where you're also trying to recruit sites into a study,

[00:15:53:050] recruit sample members into a study, and then get the data use agreements

[00:15:57:070] signed in time, so that your consent form can be used for recruiting sample members,

[00:16:01:080] there's definitely a big time crunch. So, we're hoping

[00:16:05:100] the data come in on the exact day that we're expecting it from the states

[00:16:09:130] so that we can write a final report in the studies. I think, building in enough

[00:16:13:150] time for those kinds of transactions is really the big thing that I would [inaudible]

[00:16:17:160] >>MOLLY IRWIN: So...

[00:16:21:170] thinking, sort of, to the beginning of the process, what recommendations

[00:16:25:200] do all of you have for how researchers

[00:16:29:220] can– or what they should do to do their homework

[00:16:33:260] before requesting the the administrative data. And, I know some of these

[00:16:37:280] things have already been talked about this morning

[00:16:41:290] really understanding what different administrative data are available

[00:16:46:010] what data are in a particular data set,

[00:16:50:040] things like that. What should we really be thinking about upfront, before we

[00:16:54:060] even make the request? >>JENNIFER NOYES: So, I'll jump in. >>MOLLY IRWIN: OK.

[00:16:58:070] >>JENNIFER NOYES: I think one of the big challenges we've already touched on

[00:17:02:100] multiple times, but is this issue of clarity in what you're trying to accomplish.

[00:17:06:130] Maya has brought it up, and it's come up in every presentation so far.

[00:17:10:170] And, one of the things I think researchers really need to do is think about

[00:17:14:170] the question that they're really interested in, and what it might mean when you're querying the data.

[00:17:18:190] I know that Maria this morning really talked about getting a

[00:17:22:210] thinking about it as you're sampling a population, and, um... She and I think

[00:17:26:210] a lot alike in relationship to that– that it's not just simply a matter of asking for the

[00:17:30:240] data fields and walking away. From my perspective, one of the key things– pieces of homework

[00:17:34:250] that needs to be done, is spending time with the people who actually understand the data

[00:17:38:260] to figure out what you should be asking about. Um...

[00:17:42:280] Really a big, key challenge is thinking about– talking about what you're

[00:17:47:010] trying to accomplish– what is it that you're trying to answer, so that you can go through and really

[00:17:51:040] describe the population that you're interested in the unit of analysis

[00:17:55:060] that you're interested in, the time period, your definitions of program participation–

[00:17:59:070] what it is you're actually talking about. Because a lot of times, as outsiders coming into data

[00:18:03:090] we don't necessarily understand what it represents. So, I think there has to be a big investment

[00:18:07:110] of up-front time, so that you're not wasting time after. It takes a lot of

[00:18:11:130] time to get access. And you want to make sure that the information that you're accessing

[00:18:15:140] is what you think it is, and it's going to help you answer the question. And, you might not

[00:18:19:150] need, as was mentioned this morning, everything that you think you need. We have been

[00:18:23:170] renegotiating and just finished our data sharing agreement with our Department of Corrections,

[00:18:27:190] to update it. And, one of their big hang-ups was we wanted information

[00:18:31:230] about the institution from which an offender was being released

[00:18:35:240] and it turns out that some of the institutions from which the offender is being released

[00:18:39:250] actually if we knew what that were we would know their mental health status; we would know information that's protected

[00:18:43:270] under one of the other things that was discussed this morning. And, I— and it was this huge hang-up

[00:18:47:290] with the attorneys for like two months, and I finally said, "we actually don't care. You can mask that data"

[00:18:52:000] But we hadn't understood that when we'd asked for that field that that's what we were going to get.

[00:18:56:010] And so, in my opinion, it goes back to investing the time up-front

[00:19:00:020] to talk with the people who actually own and understand the data, so that

[00:19:04:030] you know that what you're asking for is going to actually answer the question that you think

[00:19:08:050] it's going to answer, and also you don't get more than you need, because that will save you

[00:19:12:080] problems later on.

[00:19:16:120] >>BETH GREEN: I would agree with that, just as– because I was thinking the same thing, I think, when you talked about

[00:19:20:120] Think about what you really need, and think about where you can be flexible. Like...

[00:19:24:140] Here's our master list of all of the things that we're interested in looking at.

[00:19:28:170] And conceptually thinking through that research question. But then, here are the

[00:19:32:200] not– sort of– for– for the research project to even be worthwhile doing

[00:19:36:210] here's the information that we'd need to have, and be able to be flexible

[00:19:40:220] around that. Because, I think in that initial– the initial phase of

[00:19:44:240] being a researcher asking for data is really about relationship building.

[00:19:48:260] And so, to the extent that you can be very flexible and open

[00:19:52:290] and, sort of, try to have that kumbaya experience.

[00:19:57:020] I think that's really important. Keep that kumbaya experience past the first meeting,

[00:20:01:030] I should say. [others start to speak] I think the other thing

[00:20:05:050] Is that one of the pieces of homework that I didn't anticipate Would be

[00:20:09:070] as challenging as it was, was just figuring out who in the agency do I

[00:20:13:110] go to to ask. And sometimes that turns out to not be that obvious.

[00:20:17:120] And I've learned, because I'm not a person who picks up

[00:20:21:130] the phone easily, and I don't think now we do– we do most of our contacts via email,

[00:20:25:140] I've learned that the best way to get that information is by picking up the phone and calling

[00:20:29:170] somebody, and finding out who is the person I need to talk to.

[00:20:33:200] >>CHARLES MICHALOPOULOS: And I think, I'd second both the things Beth said. I think in our case, we kind of knew

[00:20:37:200] which data we wanted. We had an experienced research team that had used all the data sources that we were trying to get.

[00:20:41:210] So, we understood them—

[00:20:45:220] Not every detail in every state, but understood the basics of them and knew that we wanted them.

[00:20:49:240] But, uh...both the flexibility is really important. I think the place where

[00:20:53:260] flexibility played the biggest role in our— you know— our work was

[00:20:57:290] being willing to live with de-identified data, if that's all the state could give. In most cases,

[00:21:02:000] the vast majority were able to get identified data, so we can link them across data sources,

[00:21:06:010] do lots of interesting analyses. But we're willing to get de-identified data

[00:21:10:030] if that's all a state is willing to give us. And, it usually was because of legal restrictions.

[00:21:14:050] And then, just to echo, we also had

[00:21:18:090] the issue of who is the right person to contact at the state—

[00:21:22:100] getting the right name— do you start at the top? Do you start at the data person? — and that sort of thing

[00:21:26:110] And, um... getting those people to respond is... you know... is— is— yeah...

[00:21:30:130] probably one of the toughest things. So... I see a lot of—

[00:21:34:160] We have a weekly— we have a log of activities in doing data acquisition.

[00:21:38:190] You see lots of things like: sent an email to somebody. Two weeks later: Called them.

[00:21:42:200] Two weeks later: followed up again. And, you just see that going on, and it's not surprising

[00:21:46:220] because they have lots of other responsibilities. What we're asking them for is not

[00:21:50:240] necessarily at the top of their list of things to do. When we work directly with states

[00:21:54:260] on things that they're interested in, the data come very quickly.

[00:21:59:000] and a lot more easily. >>MAYA BERNSTEIN: I was going to add that in part of the planning for

[00:22:03:010] your work, at some point, you're going to come across somebody like me.

[00:22:07:030] And you should think about that in advance. And you should try to make it easy

[00:22:11:050] for somebody like me to say yes to what you want to do. One of the ways to do that,

[00:22:15:080] is to have thought about the whole lifecycle of the data that you're going to collect up front.

[00:22:19:100] So you want to think about how you're collecting it or gathering it,

[00:22:23:110] how you're going to use it, how you're going to manage it while you have it,

[00:22:27:120] how you're going to report out your results, and eventually

[00:22:31:140] how you're going to dispose of the data— or archive the data— at some point, what is

[00:22:35:180] the end of this. You might not know all of that up front, but it's good to have in mind the entire life cycle,

[00:22:39:200] so that someone like me knows that you planned and you thought about it.

[00:22:43:220] And, you can get help for each of those steps along the way.

[00:22:47:230] But, I think it's important to kind of have an overview before

[00:22:51:260] you start, of all the way through the project, what is going to happen next

[00:22:55:290] and where you're going to end up. >>MOLLY IRWIN: So, just to follow up,

[00:22:60:000] A lot of you focused on finding the right person.

[00:23:04:010] What's even the right kind of person? I mean I there's somebody like Maya

[00:23:08:020] I think Maria this morning talked about maybe getting in touch with

[00:23:12:050] the person who's entering the data or a case manager.

[00:23:16:080] Who— how— What's the homework

[00:23:20:100] even to do to figure out the right questions and then to find

[00:23:24:120] that person? There's the legal person. There's the privacy officer

[00:23:28:140] There's the person who knows what's actually in the database.

[00:23:32:120] >>JENNIFER NOYES: So, from our experience

[00:23:36:150] um... because we've been working with the different state

[00:23:40:160] agencies for so long and over time, when we go to bring in a new one,

[00:23:44:170] in terms of working with us, like our Department of Public Instruction, we always

[00:23:48:180] do the pitch on what's in it for them. So, we go to the program person

[00:23:52:200] first, to make the sell of "if we have access to this data

[00:23:56:230] and can do research questions, and here is the— here's the question that we're

[00:24:00:270] interested in. Is it interesting to you? And if it's not interesting to you, is there something that we can add

[00:24:04:280] to make it be interesting to you?" To develop an advocate within who will go to bat

[00:24:08:290] for us. And that has worked very, very well over time.

[00:24:13:020] We as an institution, the Institute of Research on Poverty, has also had the luxury that we have

[00:24:17:040] a lot of— we have a very symbiotic relationship with the state agencies with whom we work.

[00:24:21:060] So, I come from being a data owner and understand

[00:24:25:060] the data systems from where I came, and have a lot of connections back into the agency.

[00:24:29:080] But we didn't have a connection with our Department of Public Instruction, but we found somebody

[00:24:33:100] in their evaluation area who was very interested in using data.

[00:24:37:130] And, we figured out a very first project that we needed the data for the research we were interested in,

[00:24:41:160] and like I said we tacked on what they were interested in, and then from there

[00:24:45:180] they could tell us, "now we need to talk to our legal council. Now we need to talk to

[00:24:49:190] the, you know, basically this person's boss, and it wasn't— it was a

[00:24:53:210] blind call, but it wasn't blind in so much as we new the program

[00:24:57:240] that we wanted to work with.

[00:25:01:250] >>MOLLY IRWIN: Any thoughts on that? So, so a follow up...

[00:25:05:260] Jennifer, what are some of the hidden costs to using

[00:25:09:270] administrative data, and what can researchers do to anticipate

[00:25:13:290] or minimize those? >>JENNIFER NOYES: So, to me, the big, huge hidden cost is

[00:25:18:020] actually not so hidden, because Charles talked about it, and I was just talking to Bob about— [inaudible] about this.

[00:25:22:060] Do not underestimate the time— the time

[00:25:26:070] So, you can think about the cost of the servers, and you can think about the cost of cleaning the data,

[00:25:30:080] which I actually think is a big— is a big investment. It's not just

[00:25:34:100] pushing the button and getting the data. You have all this up front time

[00:25:38:130] to get access to the data that you need to build into any schedule that you have.

[00:25:42:140] You have the time that you need to invest in working with people like Maya.

[00:25:46:150] And the thought— There's a lot more thought that goes into this ahead of time than I think people

[00:25:50:160] realistically build into their schedules. And then, even once you have access

[00:25:54:190] I think there's a lot more time that goes into understanding if you've never used the data

[00:25:58:210] the data before— um... the data that you're going to get. And there's a hidden cost

[00:26:02:250] to the agencies with whom you're working, because you're going to be pestering them to help you understand

[00:26:06:250] what's there. And again, if you show them how it's a value to them to invest the time

[00:26:10:270] time with you, that they're going to get something out of it, you can make more progress.

[00:26:14:290] I do think that we, as researchers, can make an assumption that we know

[00:26:19:020] what's in the data. And I'm going to say, we might think we know what's in the data based on past experiences.

[00:26:23:030] But if you're working cross-states or cross-jurisdictions

[00:26:27:040] they don't have to put the data in the same way across states or across jurisdictions, not as

[00:26:27:050] they don't have to put the data in the same way across states or across jurisdictions. Not as a national

[00:26:31:070] project that we're working on that's pulling in administrative data from eight states

[00:26:35:100] and trying to get their code books, and trying to get them— for us to understand what their fields are.

[00:26:39:130] There's not consistency. So, we know our little Wisconsin child support data

[00:26:43:140] really, really well having worked on it for many, many years and now we're pulling in information

[00:26:47:150] from multiple sites, and the amount of time that we have to invest because codebooks don't

[00:26:51:170] exist, or whoever wrote the codebook is gone. So, to me

[00:26:55:190] a main hidden cost is the cost of your time as a researcher,

[00:26:59:200] and the cost of the time on the end of the agency who is helping you,

[00:27:03:240] the cost of the legal staff, and getting information to them,

[00:27:07:250] and then— we just had a case where we invested two years

[00:27:11:270] in a data sharing agreement, that came over— so I work in

[00:27:15:290] a university setting— came over and because it involves HIPPA data, and we were—

[00:27:20:020] we had to re=prove, sort of, internal to the university our HIPPA protections,

[00:27:24:040] which were in place, but the amount of time that took

[00:27:28:050] the data sharing agreement expired from the state agency—

[00:27:32:060] Well, it had to be signed within sixty days. The sixty days elapsed.

[00:27:36:080] We got everything in place to get our signatures from our legal staff, sent it back over to the state agency

[00:27:40:120] A new person has come in; now they want to review the agreement again. OK, we've been working on

[00:27:44:120] this now for two years. Luckily we built in a huge

[00:27:48:150] huge, huge lead time. So, to me, as I said,

[00:27:52:160] a big hidden cost, which isn't so hidden, but... is the time.

[00:27:56:190] >>MOLLY IRWIN: Does anybody else want [inaudible] >>BETH GREEN: Yeah. I mean, I totally

[00:28:00:230] agree with that. We have— I just got an email on Monday

[00:28:04:240] from one of the states that we've been trying to set up an agreement with for almost two years

[00:28:08:250] almost two years to the date, saying "I think we finally have someone

[00:28:12:260] who can review your request at our our legal office. So, you've moved to the head of the queue

[00:28:16:290] and this is after two years of— I mean— we have a project manager

[00:28:21:020] who I— who self-describes as a friendly pit-bull. I mean, she is on these

[00:28:25:050] places, because you have to be. If you don't keep following up, which is another time

[00:28:29:060] I mean, it's not— you're not getting the data, but you're spending a lot of time

[00:28:33:080] re-contacting and re-contacting and re-contacting and trying to be friendly and trying not to

[00:28:37:080] cross that line between being sort of a stalker and just really annoying,

[00:28:41:110] to gentle, friendly reminders

[00:28:45:120] and check-ins on, sort of, where things are moving along. So that time has been huge.

[00:28:49:130] No cost extension has been something that we've had to sak for

[00:28:53:140] many times... Because we just can't get it done

[00:28:57:170] in a short period of time. With three years being a short period, five years being somewhat longer.

[00:29:01:200] I would also say, and I think this is important

[00:29:05:200] everyone keeps speaking about the fact that these data sets are messy

[00:29:09:210] and I, as a— I was trained as a psychologist

[00:29:13:220] and, you know, as a graduate student, and my graduate students

[00:29:17:240] tend to be trained to work with these nice, clean

[00:29:21:270] flat, survey data files, that, you know, have all the nice

[00:29:26:020] documentation of exactly what variable everything is.

[00:29:30:030] That is not the case for administrative data sets. And actually finding people with the

[00:29:34:040] skill-set who can put these data sets together

[00:29:38:070] really on a technical level is harder than you would think.

[00:29:42:100] Um... We had— We just had a recent hiring experience where

[00:29:46:100] we have learned that you can't just ask interview questions when you're hiring.

[00:29:50:110] someone to do this kind of work, you actually have to give them a task, and ask them to do it.

[00:29:54:130] So, give them a data set and ask them to reconfigure it in certain ways

[00:29:58:150] and to look for these things in the data. We had several experienced analysts

[00:30:02:180] who basically could not even start to tackle that problem.

[00:30:06:200] And we'd have never known that in the hiring process if we hadn't have asked them to

[00:30:10:210] to do something hands-on. So, that's been a lesson learned. I think it's important.

[00:30:14:220] >>CHARLES MICHALOPOULOS: I just wanted also to say, it's important not to conflate, because I've heard both

[00:30:18:240] mentioned here, the question of which data sources from what's in them

[00:30:22:280] and how do you use the data source. So, for example, one of the questions

[00:30:26:280] in one of our home-visiting studies in the effect of home-visiting on Medicaid costs and Medicaid use.

[00:30:30:290] So, it's clear Medicaid data on claims and encounters are going to be a good source of that information.

[00:30:35:010] But, each state, you know, um, puts their Medicaid in different forms.

[00:30:39:020] And so, understanding exactly what is in each state system is taking a lot of time.

[00:30:43:050] Even though we knew that was the data set we wanted to get in each state.

[00:30:47:080] >>BETH GREEN: Well, and for example, these things change over time. We used to— we did a lot of work in the state of Oregon

[00:30:51:100] with their addictions and mental health data sets, and it used to be

[00:30:55:120] quite a great source of data— fairly accurate. It was used as a billing system.

[00:30:59:140] So, people— treatment providers— who wanted to get paid actually put their data in.

[00:31:03:170] And then a few years ago, for some reason I don't understand

[00:31:07:190] it is no longer a billing system, and the data, even the state

[00:31:11:200] research agency folks will say, "Mmmm... I'm not sure you really want to

[00:31:15:210] trust this data if you're looking at something like how did this program

[00:31:19:230] increase the rate of access to treatment services, for example.

[00:31:23:260] Because half of the treatment providers we work with don't data into the system."

[00:31:27:260] >>JENNIFER NOYES: So, can I jump in on sort of this theme?

[00:31:31:280] One of the things that we have definitely learned, and we've talked about integrating data across systems,

[00:31:35:300] is that different data systems privilege different

[00:31:40:020] different data. And, that you can— If it's important to that program, that set of data

[00:31:44:040] is going to be more reliable and easier to use.

[00:31:48:080] like Medicaid claims. And, if it's information that is nice to be

[00:31:52:100] inputted for case management or other administrative purposes, but it doesn't really matter,

[00:31:56:120] then— so there's this hierarchy across all the data systems

[00:32:00:140] about which fields are more or less believable. You know, sort of like the—

[00:32:04:170] I don't know what the right word would be. But, we have learned working across

[00:32:08:190] all these different systems, in terms of going to the probabilistic matching that we're using

[00:32:12:200] to be able to create a flat file that's one person across time, and then connect to their

[00:32:16:210] case or to their parent or whatever it is, is that there—

[00:32:20:240] as a third party sort of arbiter, to try to bring the systems together

[00:32:24:270] one system might be– won't want to have their information necessarily overridden, but–

[00:32:28:280] If you can get a Social Security number from UI wage records, then it's probably much more valid

[00:32:32:290] than anybody who would have picked a Social Security number up in a Chile Welfare record, right?

[00:32:37:010] And so, you learn that, over time different things are more or less

[00:32:41:020] going to be valid. And if it's really important to the program, like claims data

[00:32:45:040] or utilization data, in the medical area.

[00:32:49:040] Or familial relationships, maybe, in a Child Welfare data set,

[00:32:53:070] then we chose to privilege that information in a data set and ignore, maybe,

[00:32:57:080] if there's conflicting information in a different data set. And in my mind, that's the only way

[00:33:01:100] you can bring things together. And again, that's a hidden cost, because you have to learn

[00:33:05:130] enough about the program, or be working with people who know about the program

[00:33:09:160] to get them to tell you what is important to them or not. Right?

[00:33:13:170] And, just because it's there, doesn't mean it's true. [laughter]

[00:33:17:180] >>MOLLY IRWIN: Great. So let's

[00:33:21:200] So, let's turn now to gaining access

[00:33:25:230] to different types or levels

[00:33:29:270] of data. So, what are the unique challenges

[00:33:33:280] of gaining access to federal versus state versus county

[00:33:37:290] or local level data? And do the levels require

[00:33:42:020] negotiating with different— with each of those levels

[00:33:46:050] require negotiating different agencies or organizations?

[00:33:50:080] Are the laws different? Regulations? Et cetera...

[00:33:54:100] [multiple people mumble]

[00:33:58:110] [laughter] >>CHARLES MICHALOPOULOS: Yeah... all of the above. I mean we've heard

[00:34:02:120] a lot already about the challenges of getting data from states.

[00:34:06:160] From the earlier session and from today. And they're all true. and if you're working with three

[00:34:10:170] state agencies, then just triple all that; if not more, because sometimes

[00:34:14:180] the state agencies don't, you know— I think we heard earlier— sometimes they don't want to share data.

[00:34:18:200] Possibly because they're worried about misuse It was easier for us

[00:34:22:220] because we never had the intention of sharing them with other state agencies;

[00:34:26:260] it was always to assemble them for the research and collect them

[00:34:30:260] just for the purpose of the research. So, I don't think that was as big a problem for us. But, I think it's just taking all

[00:34:34:270] all the things we talked about and multiplying it across forty state agencies

[00:34:38:290] that we were trying to work with for the studies.

[00:34:43:020] So, in that sense if you can find a federal, national database

[00:34:47:050] like the New Hires data, or you can do your work with the SSA data or

[00:34:51:060] um.. you know... um... disability data

[00:34:55:080] that sort of thing, then you should definitely do it. There are other restrictions

[00:34:59:090] for new hires you have to be— you have to be

[00:35:03:120] showing them that you're providing some information of interest to the agency.

[00:35:07:150] that has the data, you know— the Office of Child Support Enforcement. So, not anybody can use them,

[00:35:11:170] but if you can do that, then it's good. In our case, we were working with

[00:35:15:180] with folks at ACF, so they were able to actually

[00:35:19:200] do all of the negotiation. So, it was a lot easier for us to get access to the

[00:35:23:220] New Hires data, because somebody else did all the negotiation within the same agency,

[00:35:27:240] than it was to get it from the state. But, I think, um...

[00:35:31:250] all the challenges we talked about for state agencies apply, and the key is,

[00:35:35:260] I think, being flexible. We talked about who should you be contacting. And the answer again, is all of the above.

[00:35:39:290] Like anybody who might be helpful. Sometimes we start it

[00:35:44:010] with the data person, and they were great and they helped work our way up to the right people.

[00:35:48:040] Sometimes that didn't work, and in one case we talked to

[00:35:52:060] the Medicaid medical director, and they were able to

[00:35:56:070] help us figure out the right person to get things. So, it's really just being flexible

[00:36:00:100] and trying everything that you can.

[00:36:04:130] >>JENNIFER NOYES: So... I'll just keep echoing

[00:36:08:170] from our experience in terms of building relationships

[00:36:12:170] is, as I said, trying to— I think our success in gaining access

[00:36:16:190] has really been, on a state level, to think of this as an over time,

[00:36:20:220] a long term, a horizon. And, to figure out a way to

[00:36:24:250] at least demonstrate initially from maybe a limited

[00:36:28:260] access to the data, to show that we can be responsible

[00:36:32:260] and develop research that is of a interest to whoever provided the data. Then to build that over time.

[00:36:36:270] To the point where our goal with every agency with whom we work

[00:36:40:290] is to have more of a master data sharing agreement that is general around

[00:36:45:020] the provision of the entire data set, and then as we have specific projects, we add it as addendums

[00:36:49:030] or attachments. One of the ways that has been very successful lately

[00:36:53:040] because everything that we do is crossing state agencies, is using integrated data—

[00:36:57:060] Excuse me— data, is that when we're working on a research project that involves multiple

[00:37:01:080] state agencies that we want to get permission to use their data, is we pull together meetings

[00:37:05:100] with all the actors together, instead of going one by one by one.

[00:37:09:140] Because the biggest challenge will be, perhaps we have a question

[00:37:13:160] of interest to the Department of Public Instruction that is going to involve using data

[00:37:17:180] from our Department of Children and Families. And instead of going to pitch it to the Department of Children and Families,

[00:37:21:190] separate, we get the people in the room together to show who's of utility across all of their

[00:37:25:210] programs. And so, even though we have separate data sharing agreements with each agency

[00:37:29:240] we at least get buy-in managerially or at the

[00:37:33:250] top level that this is what wants to be supported.

[00:37:37:270] So for us, within the state it definitely been a long term investment.

[00:37:41:280] And again, building on— I think it was mentioned earlier— get your foot in the door and

build from there.

[00:37:46:010] And— and— that's what we have been able to do.

[00:37:50:050] But, I'll sound like a broken record— if people talk about broken records anymore— about this

[00:37:54:060] whole relationship building, and that, in my mind, to gain access and to

[00:37:58:080] maintain it, you need to show what's in it for other people and not just for yourself as a researcher.

[00:38:02:100] >>BETH GREEN: And, sometimes that's more challenging

[00:38:06:130] like I think of our situation, where in our early Head Start study we were trying to go back

[00:38:10:150] and look at this very large sample. We collected

[00:38:14:160] for that original study, lots and lots of survey

[00:38:18:170] assess— child assessment data— and actually I see Luisa back there

[00:38:22:200] she knows a lot about the early Head Start study [laughter]

[00:38:26:220] And then, look and see— does this early childhood program

[00:38:30:240] really have an impact on child maltreatment outcomes. and so we're working with

[00:38:34:250] child welfare agencies, who maybe have sort of a general

[00:38:38:260] interest in child abuse prevention and building that evidence base

[00:38:42:280] but there's nothing really really in it for them. I mean, we tried as hard as

[00:38:47:010] we could— I mean that we said we'd share back their data, you know, we would do whatever ever we could

[00:38:51:020] to help them understand the findings. But really, they have to say

[00:38:55:030] "OK, we agree that this is an important

[00:38:59:040] potential contributor to the evidence base around child abuse prevention."

[00:39:03:070] That said, I will tell you that we did run into one agency,

[00:39:07:100] because this was actually part of a different study with child welfare, where we were trying

[00:39:11:140] to do a cost-benefit analysis of child abuse prevention

[00:39:15:140] and we wanted access to cost data from the child welfare agency, and the person said

[00:39:19:160] "So, if your results are positive, you're going to basically

[00:39:23:180] have data that helps provide a rationale for cutting

[00:39:27:210] our state agency staff." So... [laughs] We thought it was great—

[00:39:31:220] child abuse prevention is a good thing, everybody wants that, and I'm sure this person did too,

[00:39:35:230] but when push came to shove, she said, "aren't you basically

[00:39:39:240] saying that child welfare agencies are spending more than they should?

[00:39:43:260] and we could actually be putting the money in prevention?" And in a state budget, which is tight,

[00:39:47:290] maybe not such a positive thing. So, getting that buy-in can be more

[00:39:52:030] more or less difficult. >>CHARLES MICHALOPOULOS: Can I just mention too

[00:39:56:040] We talked about a buy-in, but part of— just another example of being flexible...

[00:40:00:060] We had one state agency that didn't want to know the results of our evaluation.

[00:40:04:090] So, that makes it a little bit more challenging to get buy-in. Like, "We can tell you

[00:40:08:120] great things that are happening." [laughter] Because, they had a program that they thought was effective.

[00:40:12:130] It was getting funded by the state, and they didn't want results out there that would suggest the opposite,

[00:40:16:140] similar to, I think, what Beth was talking about. So, that was a case where we promised,

[00:40:20:160] "Don't worry, we're not going to reveal your state's results to anybody. Nobody will ever

know

[00:40:24:180] whether it was effective where it was"

[00:40:28:210] That was already our intention anyway. We had to go one step further and say, "When we release it

[00:40:32:230] public use file for this, we'll make sure that nobody will be able to identify your state

[00:40:36:240] and be able to produce state results."

[00:40:40:260] >>MOLLY IRWIN: So, it sounds like many of the things that you guys are talking about are— are

[00:40:44:280] the same across different levels. Is there anything

[00:40:49:020] that comes to mind that's different if you're working at the local level

[00:40:53:030] versus the federal level, in terms of laws, regulations,

[00:40:57:040] the folks you would be talking to, ho you get

[00:41:01:050] in the door, the homework you would be doing Or are these

[00:41:05:080] these... these... strategies and issues pretty consistent?

[00:41:09:110] across levels. >>BETH GREEN: I can't speak to the federal issues

[00:41:13:130] because we've never tried to access a federal database, but we've accessed state and county

[00:41:17:140] databases, and I— they feel the same, to me, in terms of

[00:41:21:160] the proc— I mean, every state is completely different. So, each one

[00:41:25:180] has very a very different process and different hoops you have to go through,

[00:41:29:190] and different kinds of barriers too, that— I learned... I thought once we got a few states to agree

[00:41:33:200] to come on board, it would be easy to convince the other ones. We were working in a situation

[00:41:37:220] for the early Head Start study, where we— kind of a difficult

[00:41:41:230] combination of factors, where we wanted identifiable, leveled

[00:41:45:260] child-level data. We did not have consent, so we were operating through a

[00:41:49:290] waiver of authorization of informed consent. And we wanted this historical data.

[00:41:53:290] But once we got a few states to agree, I thought it would be

[00:41:57:300] a piece of cake. But, no; it was not. I will say that some states

[00:42:02:010] that you think would have a state agency database, do not.

[00:42:06:040] So, for example, Pennsylvania does not have a state child welfare

[00:42:10:060] database. [laughs] So, they have ninety-something counties in the state

[00:42:14:100] of Pennsylvania, so if we wanted to get data that was comparable to our other states,

[00:42:18:100] we would have had to go to ninety different counties and negotiate ninety different

[00:42:22:120] data sharing agreements, which obviously was not in our budget to do that. So...

[00:42:26:140] That was an interesting lesson learned. >>CHARLES MICHALOPOULOS: And, I think one

[00:42:30:170] difference is

[00:42:34:190] Um... You know... At least for the federal databases that I'm familiar with

[00:42:38:190] they're— you know— the agency that has them

[00:42:42:200] is familiar with research, has worked with researchers, and so the issues around

[00:42:46:220] how to use the data for research are pretty clear. We definitely talked to state agencies

[00:42:50:240] where no one had ever asked them for the data that we were asking for for research

[00:42:54:250] purposes

[00:42:58:260] At least for the kind of study we were doing. So, they weren't even sure how to proceed,

[00:42:58:260] who to talk to

[00:42:58:260] what the legal restrictions were, and so on. So, I think that's definitely

[00:42:58:260] a pretty substantial difference, that makes it much easier

[00:43:02:270] in a lot of ways to use federal data, even though there are

[00:43:06:280] some— some drawbacks with those as well. >>JENNIFER NOYES: And, I think

[00:43:11:010] if you talk about local data that it becomes even more the case, that you don't necessarily

[00:43:15:040] have the capacity locally. People, um...

[00:43:19:060] will be a lot more conservative either in terms of understanding— well, it will go one of two ways

[00:43:23:070] Either it will be super loose, because they don't have that legal person who's telling them that they can't give it to you,

[00:43:27:080] or they'll be super conservative, because they're not quite sure if they can.

[00:43:31:100] And even if they want to be cooperative, and work

[00:43:35:150] with you to access the data, they won't necessarily have the capacity to

[00:43:40:020] actually help you that done. They won't know quite how to get it, like local jail data

[00:43:44:050] that type of thing. They don't have some just to task— give a task order to,

[00:43:48:060] or something like that... >>MOLLY IRWIN: So— [multiple people speak]

[00:43:52:070] >>MAYA BERNSTEIN: You did ask about legal and regulatory stuff. [inaudible speech] Go. Go ahead.

[00:43:56:090] >>MOLLY IRWIN: So, I was going to— >>MAYA BERNSTEIN: Multiple conversations.

>>MOLLY IRWIN: Well, I was going to

[00:44:00:110] and I bet this is where you're going— I was going to turn to you and ask who the various stakeholders

[00:44:04:120] in the decisions are about accessing administrative data.

[00:44:08:160] What can be shared? How the data can be used?

[00:44:12:170] >>MAYA BERNSTEIN: OK. Uh... You know...

[00:44:16:190] Your most important stakeholder is your program people, right? They're the people who

own the data.

[00:44:20:220] And their— you know— their first priority is their program; it's not your research

[00:44:24:250] They still have to run their program. And, so, when you talk about

[00:44:28:270] taking time away, you know, cajoling them, and all that

[00:44:32:280] they're working on managing the program and trying

[00:44:37:000] to fit in your, perhaps, ad hoc request for research

[00:44:41:010] in the middle of that. And so, I do

[00:44:45:030] quite a lot of work with OCSE on the National Directory of New Hires,

[00:44:49:040] and I know that they are very concerned about being, you know, true to

[00:44:53:050] their main mission, which is child support enforcement.

[00:44:57:070] That's what the database was created for. And so even though we have

[00:45:01:090] statutory requirements to share data with various other programs,

[00:45:05:110] for other administrative purposes...

[00:45:09:130] it's still— their priority is protecting the data for the original

[00:45:13:140] purpose for which it was collected, which is child support enforcement. And I think, all of the programs

[00:45:17:150] at whatever level of government that you are looking at are going to have that

[00:45:21:170] same view of the world. So, that's one thing...

[00:45:25:180] At the federal level, which is where my expertise is

[00:45:29:210] I can't help but point out that Sunday...

[00:45:33:230] was the fortieth anniversary of the implementation of the Privacy Act of 1974

[00:45:37:240] you might how that is, since it's 2015 and not 2014 [laughter]

[00:45:41:260] The law was signed by President Ford on December 31st

[00:45:45:280] on New Year's Eve 1974, but the implementation date was the following September

[00:45:50:020] And so it's forty years Sunday. And most all

[00:45:54:030] the databases that we're talking about at the federal level are going to be covered by the federal privacy act.

[00:45:58:050] The Federal Privacy Act is not that private. But, it does have

[00:46:02:080] a statutory provision that says you may share the data for research if it

[00:46:06:100] in a not identifiable form. That is just not that useful, it turns out.

[00:46:10:110] Right? Most of the time when you're doing your research you want identifiable data

[00:46:14:120] at least to create your initial database for doing your research.

[00:46:18:130] You want to link it to something else, and without those identifiers you can't link it.

[00:46:22:150] So, that's not that useful, and we have other ways in the Privacy Act of

[00:46:26:170] allowing the data to be used if it's compatible

[00:46:30:180] with the reason for which the information was collected. And, compatibility is not defined

[00:46:34:190] in the act, so we have a lot of wiggle room there. I should say, by the way,

[00:46:38:220] I am not legal council at my agency. I'm the policy arm

[00:46:42:230] of the secretary, even though I am a lawyer by training, but I'm not practicing law at the department.

[00:46:46:260] And so, that might be another hint, that there might be somebody

[00:46:50:290] else who's not the legal council who can help you.

[00:46:55:010] Whose job it is to think about policy matters and what makes sense

[00:46:59:020] for the agency and its programs. Rather than the legal council, who is...

[00:47:03:040] let's say... a little more constrained... perhaps...

[00:47:07:070] The other big law that you're probably dealing with a lot— people have mentioned

[00:47:11:100] HIPPA already, and you talked about whether

[00:47:15:130] different agencies are covered by different laws in different

[00:47:19:150] state levels— That's absolutely true. So for example,

[00:47:23:170] if you have data that's coming from the federal Medicaid or Medicare program

[00:47:27:200] they're a HIPPA— CMS is a HIPPA-covered entity. It's covered by HIPPA. It's also

[00:47:31:210] It's also covered by the Privacy Act. But, if you move— If the data moves

[00:47:35:230] to another part— even to move to ACF; It's no longer covered by HIPPA

[00:47:39:260] because ACF is not a HIPPA-covered entity. It's still covered by the Privacy Act.

[00:47:43:270] If that data moves to a state... OK... it might be— you know— it might be

[00:47:48:000] covered by HIPPA, because it's a covered entity.

[00:47:52:010] It's the public health agency, for example, And they're actually billing electronic claims.

[00:47:56:020] or they're actually a provider or a plan they're a Medicaid agency, for example

[00:48:00:040] But it's not covered by the Privacy Act, which only covers federal agencies, right?
[laughs]

[00:48:04:060] With respect to the privacy of data, the rules

[00:48:08:070] do not follow the data around; they follow the custodian of the data.

[00:48:12:110] And so, for each custodian of the data, you're going to have a little bit different

[00:48:16:110] set of rules. And you have to find the person who's the custodian of that data

[00:48:20:120] to help you understand what the rules are with respect to that data set.

[00:48:24:140] And so, when you're matching things up you're going to maybe come across

[00:48:28:170] different kinds of privacy rules. There are also ones, for example OCSE,

[00:48:32:200] has its own rules under the Welfare Reform law

[00:48:36:210] from '96 that specifically protect that database, and say it shall not be used.

[00:48:40:220] Right? For anything except— That's the typical structure of a privacy law.

[00:48:44:250] And in the case of the National Directory of New Hires, it's 4D program

[00:48:48:280] it's Social Security, and it's Earned Income Tax Credit. And then since then,

[00:48:52:290] we've had statutory exception that add new uses.

[00:48:56:290] But, for that data you need a statutory exception every time.

[00:49:01:010] If you're not in those, you're not getting the data. You're one of those. That's good
[laughter]

[00:49:05:030] [laughter] You have state laws. A wide variety of state laws.

[00:49:09:060] that protect different kinds of data. Some states protect

[00:49:13:100] things that you understand why they were created.

[00:49:17:110] So, we have some states that protect particularly careful

[00:49:21:120] sickle-cell data in medical records— separate differently

[00:49:25:150] You have lots of HIV and AIDS, kind of,

[00:49:29:170] protections that popped up in the '80's and '90's, but not all

[00:49:33:190] the states have the same one. You can even find states that have protections on

[00:49:37:200] ophthalmologic data— just, other kinds of bizarre things. So, you have to talk to

[00:49:41:210] the program person who knows their data to understand

[00:49:45:230] what the rules are with respect to that data, and try to find

[00:49:49:260] a creative, you know, policy wonk or lawyer that's going to

[00:49:53:270] help you understand what the rules are, and how you may fall into an exception

[00:49:57:270] so that you can use the data for your research.

[00:50:01:290] I guess I want to mention the common rule. So many of you know that

[00:50:06:010] the common rule is now— there's been a new proposal that just came out,

[00:50:10:040] at the beginning of the month... last month... [laughs] as of today.

[00:50:14:070] And so the proposal there

[00:50:18:080] I mean I urge everyone to read it, comment on it. The comments are due

[00:50:22:090] at the beginning of December, I believe on December 2nd.

[00:50:26:110] I don't know how much it was mentioned this morning. >>MOLLY IRWIN: A little bit. It came up.

[00:50:30:140] >>MAYA BERNSTEIN: So, and the idea behind the proposal is actually to make things

[00:50:34:170] a little bit less burdensome for researchers— lots of different kinds of

[00:50:38:190] things have come up in the years— thirty years, forty years—

[00:50:42:210] since the Common Rule has been in place

[00:50:46:230] that changed the landscape of how research is done— multi-site kinds of things

[00:50:50:260] and different kinds of data analysis. And so, the rule is

[00:50:54:270] really intended to open things a little bit more for researchers

[00:50:58:280] So, I think you'll find that, with respect to administrative data, its going to change a little

[00:51:02:300] And you should look particularly at that, if that's something that you do regularly.

[00:51:07:020] Which, clearly this group does. I guess the other thing I would just add

[00:51:11:050] is that at each of these levels you're going to find different levels of sophistication

[00:51:15:050] of the people that you're dealing with. And so, um..

[00:51:19:060] you might take a— you might take into account

[00:51:23:080] that there's more or less experience in the person that you're dealing with.

[00:51:27:090] And, so, for example, the Privacy Act, it's been around for forty years.

[00:51:31:120] HIPPA– the implementation date is not– you know, it's '80–

[00:51:35:150] or whatever it was. We passed it in '96, so in 2003.

[00:51:39:160] There's much less experience with HIPPA. And, you'll find that at the beginning

[00:51:43:180] of the implementation of a law you have, what I would call

[00:51:47:200] over zealous lawyers, who are very– want to protect

[00:51:51:220] the data in ways it actually is not required by law to be protected.

[00:51:55:260] And so, you maybe can push a little and negotiate a bit and try to find out

[00:51:59:270] if there are ways to get access that

[00:52:03:280] um... that aren't– aren't– you know...

[00:52:08:000] let's say... It's helpful to find a coalition or more than one voice.

[00:52:12:020] at the table. And I think your approach is great, to get more than–

[00:52:16:050] the group of people at the table. Because, you sometimes

[00:52:20:060] find someone who is overly conservative about what can actually

[00:52:24:080] be done with the data, and that's not really helpful

[00:52:28:100] for the implementation of the original purpose of the law or for the kind of work

[00:52:32:130] that we're trying to do here.

[00:52:36:150] >>MOLLY IRWIN: Are there other stakeholders we should be thinking about

[00:52:40:170] that are part of the decision for gaining access

[00:52:44:190] like IRB's for example...

[00:52:48:210] Um... The people who are actually giving the data– Maya touched on that early on.

[00:52:52:240] Are there other things that researchers should have in mind

[00:52:56:260] when they're thinking through this decision process at the beginning?

[00:53:00:270] We've touched on many of them.

[00:53:04:280] OK... >>CHARLES MICHALOPOULOS: I mean, those things

[00:53:09:010] that the IRB isn't [inaudible] are important– >>MAYA BERNSTEIN: [talking simultaneously] IRB's, lawyers, policy people

[00:53:13:040] security people that you're going to have for your data use agreement

[00:53:17:060] yeah... you know... you're... yeah... we've hit on many of them.

[00:53:21:070] >>CHARLES MICHALOPOULOS: I was just going to comment too that there's a little bit of a conflict

[00:53:25:080] between the time constraints and the desire to have multiple people you know, talk to all the stakeholders

[00:53:29:110] and have a– scheduling that meeting [laughter] means it's going to be adding six months

[00:53:33:140] to the timeframe for doing something. So that is a– it's a tension.

[00:53:37:150] It's important to do it, but how do you do it and still do things in a timely fashion?

[00:53:41:160] It's tough sometimes. >>JENNIFER NOYES: The only other comment that I would make is

[00:53:45:170] over the time period that I have been working in this area, in terms of the data

[00:53:49:190] governance boards within a state agency, is when we first started

[00:53:53:220] working in this area twenty years ago, no agency had a

[00:53:57:240] a data governance board. And, now, almost all of them do.

[00:54:01:250] Where they have their form where you have to submit your research request,

[00:54:05:260] and that type of a thing. And so, it was mentioned before, that there's been a lot of change

[00:54:09:280] over time, about how we get this done. So, getting to know that

[00:54:14:010] governance structure within the agency is part of the homework ahead of time.

[00:54:18:030] And you can know that. That can be known– How they govern their data.

[00:54:22:040] If it's about finding a person or if it's about finding a committee or a board.

[00:54:26:050] >>MOLLY IRWIN: So... So... Let's talk a little bit more

[00:54:30:080] about sticking points in actually creating data sharing agreements

[00:54:34:100] Um...

[00:54:38:110] Who are some of the people, and what are some of the strategies

[00:54:42:120] that can be used with sticking points that come up

[00:54:46:140] with negotiating legal terms

[00:54:50:160] IRB's, the idea of who owns the data,

[00:54:54:190] and then also, more broadly, and I think this has come up already,

[00:54:58:230] the idea of how

[00:55:02:240] laws like HIPPA and FERPA are often really

[00:55:06:260] interpreted differently by different people. And, often times–

[00:55:10:280] and I think we heard this already this morning– those laws are maybe

[00:55:15:010] not often nearly as restrictive as

[00:55:19:020] the interpretation on the ground is.

[00:55:23:030] >>CHARLES MICHALOPOULOS: Well, we hit many

[00:55:27:040] sticking points... [laughter] I'll just mention a few

[00:55:31:070] One was– you know– a lot of state agencies

[00:55:35:090] want to review things before they're published.

[00:55:39:100] And, our agreement with the federal government is that nobody can review things before they're published

[00:55:43:100] in an effort to maintain the independence of the

[00:55:47:120] evaluation, it's important to not have people think that

[00:55:51:140] state agencies can influence what's in it or the program of study can influence it.

[00:55:55:160] So, it was important not to allow that to happen. So, that was definitely a sticking point that...

[00:55:59:200] Um... was... Often, we— they— I think the agency just gave in.

[00:56:03:200] But OPRE was willing to say, let a state agency

[00:56:07:220] see something like the day before it was released, so they at least know what it was in one case the state was mostly

[00:56:11:230] concerned about that we would use the data correctly and wanted to see the report to make sure we were doing that

[00:56:15:260] So, we agreed to set up regular meetings with them where we talk about what we're doing and really go through the process.

[00:56:19:290] And that sort of thing, and that satisfied them. The other big things

[00:56:24:020] I would say are about who would have access to the data.

[00:56:28:040] So we are planning to create a public use file, for example. We call it a restricted access file,

[00:56:32:060] because it won't be available to all the public, but there'll be some restrictions on who can use it

[00:56:36:080] And that was a sticking point to a lot of state agencies, who are worried about

[00:56:40:090] confidentiality and all the issues that we've talked about.

[00:56:44:100] So, we have been working through exactly

[00:56:48:040] what will go into this restricted access file to make sure that the state

[00:56:52:060] agency's needs are met.

[00:56:56:080] I talked earlier about not being able to identify particular states

[00:57:00:100] or even maybe particular local areas in the data

[00:57:04:140] So making sure that it's not just de-identified in kind of the normal ways, but

[00:57:08:160] even de-identifying states, in this example. We also— there's also

[00:57:12:170] that provision in our contract with the federal government that they may decide

[00:57:16:190] they don't like the work we're doing and in the future, somebody else will be hired to do longer term

[00:57:20:220] follow-up with families. So, that's in our consent form, and that's something

[00:57:24:250] we've raised with all of the state agencies and that's obviously

[00:57:28:260] there's concern about, again, confidentiality issues and who's going to have access to the data

[00:57:32:280] and usually the solution to that is to say, "If that happens, we're going to do our best

[00:57:37:000] to do great work, and probably that won't happen in the future, but

[00:57:41:030] if that happens, you get to— you know— whoever does it will have to reach an agreement

[00:57:45:070] with you and agree to the same confidentiality restrictions that we're agreeing to

[00:57:49:080] and that was something that satisfied— um, satisfied people.

[00:57:53:100] So, those are like some of the big things that I think came up, in addition to some of the others

[00:57:57:130] we talked about— the ones we talked about like not having the resources to provide the data.

[00:58:01:160] I think the ones that I just mentioned though, are ones where the legal counsel really came in, in addition to

[00:58:05:170] just thinking about the legalities of it.

[00:58:09:180] So, I would our— my perspective, having our own legal counsel

[00:58:13:190] was one of the biggest— one of the most important things in this whole data acquisition process.

[00:58:17:220] So, once the lawyer got involved, we stepped back and let the lawyers talk to each other.

[00:58:21:230] And, they can talk the same— a similar language.

[00:58:25:240] >>BETH GREEN: I was going to— I was just thinking while you were talking, that I think our biggest

[00:58:29:250] sticking point has actually been when the lawyers don't agree.

[00:58:33:260] Um... And, the first— the first round of

[00:58:37:280] this study that I did, I was working for a very— very small, private, research

[00:58:42:000] company, and they basically— they didn't have legal counsel, and they kind of signed

[00:58:46:040] whatever the state wanted them to, in terms of data sharing agreements, which— for better or worse.

[00:58:50:050] And then, I moved to the university setting, where the lawyers look over

[00:58:54:070] all of the contracts, and, for example—

[00:58:58:090] apologies to the lawyers in the room. [laughs] I never thought I'd be spending

[00:59:02:120] so much time talking to lawyers and talking about contracts as a researcher.

[00:59:06:140] But, we have a— basically a huge sticking point

[00:59:10:150] right now, around people in New York. Their contract—

[00:59:14:160] Their legal office insists on having language saying

[00:59:18:180] that whoever signs this will follow all of the related New York— there's a line that says

[00:59:22:230] "all of the related New York statutes governing data security and confidentiality."

[00:59:26:230] Our office in Portland State will not sign that unless they detail.

[00:59:30:260] And, they can't— they said, "We can't spend the resources detailing all of the New York statutes—

[00:59:34:270] statutes that we want you to follow." Well they say, "Well how can we follow them, unless

[00:59:38:280] we know what they are?" And they say, "Well, you guys need to go research them and

figure out what they are."

[00:59:43:000] And our lawyers say, "Well, we don't have the resources to do that." So we're essentially at loggerheads

[00:59:47:040] right now on this data sharing agreement, because of the legal stuff.

[00:59:51:050] Drives me crazy. Sorry. [laughs]

[00:59:55:070] >>JENNIFER NOYES: So, I don't have that much more to add, except for sometimes, I end up

[00:59:59:090] telling people that I am not a lawyer, I just play one on TV. Because, in order to solve some of these

[01:00:03:130] problems, we all have to get conversant in some of these areas.

[01:00:07:150] So, we're not attorneys, but we need to have an understanding,

[01:00:11:160] particularly around the different state laws, in my opinion.

[01:00:15:180] When we're doing a federal study, or when we're doing something within the state, one of the tactics that has worked

[01:00:19:200] very, very well for us lately, is when we've had issues with

[01:00:23:240] attorneys within particular departments within a state. If we have a relationship with

[01:00:27:260] a different department, we have asked— and it has worked quite

[01:00:31:270] successfully— for the attorneys in that department to— they're not even involved,

[01:00:35:280] but to talk to the attorneys in the other state agency about how this can be done

[01:00:40:010] and why certain things can happen.

[01:00:44:050] And so, what we've done— again it's the over time— if we've had a successful relationship

[01:00:48:060] with one particular attorney, or with one general counsel

[01:00:52:070] they have actually voluntarily said, "Let me pick up the phone and

[01:00:56:090] talk to the attorney in the other state agency to talk to them about how to

[01:01:00:120] overcome these barriers." And it has worked very well, and we have been

[01:01:04:130] very, very good about being very thankful for the assistance we were given—

[01:01:08:130] not by an email, but by a meeting, or I'll pick up the phone, as has been mentioned before.

[01:01:12:150] Because, sometimes the attorneys— we have our attorneys at the university, but then the attorneys within a state

[01:01:16:170] agency trust more the attorneys at another state agency,

[01:01:20:190] and so that's worked well for us. But, you can't avoid these issues. That's for sure.

[01:01:24:220] >>MAYA BERNSTEIN: I want to highlight something that you just said, which I come across often, which is that

[01:01:28:230] you asked the right question, which is, "How can we do this?" That's the question you want to ask your lawyer,

[01:01:32:240] not "if we can do this," right? Because, they are tempted to say no.

[01:01:36:260] But, if you ask, "How can we do what it is we are trying to do within the law?"

[01:01:40:280] That's the question you want to ask the lawyers. I'm not sure it will fix your problem. [laughs] >>BETH GREEN: Yeah, we tried that.

[01:01:45:010] >>MAYA BERNSTEIN: Yeah, and, um... I realized in the conversation that I

[01:01:49:020] missed my opportunity when you had that little, previous segment, but I'm going to try to connect it

[01:01:53:030] to this one. Which is, I staff an advisory committee

[01:01:57:050] to the Secretary of the National Committee on Vital and Health Statistics

[01:02:01:080] that some of you may be familiar with. And, out of the Privacy,

[01:02:05:100] Confidentiality, and Security Committee there, we worked with our Population Sub-Committee

[01:02:09:110] to put out a— what we called a tool kit for communities

[01:02:13:120] using health data. And even though it— So—

[01:02:17:140] I would hold up a nice glossy one, but it's published electronically, and I will—

[01:02:21:160] however the right way to do it is— send a link to whoever the participants are here,

[01:02:25:190] so that you can look at it. So, the tool kit

[01:02:29:200] is meant for an audience which is probably less sophisticated than the people

[01:02:33:210] who are in this room, but could still be helpful, I think.

[01:02:37:220] And the idea is for people doing work

[01:02:41:240] with localities at various levels— so, it could be neighborhood,

[01:02:45:270] it could be county, it could be states. How they could

[01:02:49:280] use and connect health data to do various kinds of research.

[01:02:53:300] And, it talks about a variety

[01:02:58:010] of ways to do that. A variety— a variety of tools

[01:03:02:030] in the tool kit. One of them is, certainly, creating relationships,

[01:03:06:060] in the way that we've been talking about here. So, I want to highlight—

[01:03:10:100] highlight that. And, another is with respect to privacy, and confidentiality,

[01:03:14:110] security, that sort of thing. It goes over

[01:03:18:140] what we in the privacy community call fair information principles

[01:03:22:160] When we think about using data, we think about a group of principles that

[01:03:26:190] underlie all of the work that we do. Privacy, for me, is not

[01:03:30:200] just "can you get access to data?" But it's a— it's a

[01:03:34:210] a group of practices, that have really to do

[01:03:38:230] with fairness. So, it's, "Did you give the person notice when you were

[01:03:42:250] collecting the information? Are you only collecting the minimum

[01:03:46:280] amount that you need to collect to do the job? Have you defined

[01:03:50:290] carefully what that job is? What the purpose of your collection or gathering is?

[01:03:54:300] And here, we're essentially—

[01:03:59:010] I would say— a little bit violating or stretching

[01:04:03:030] one of the major fair information principles. Which is, you shouldn't use information

[01:04:07:060] collected for one purpose, without going back to get the consent of the original subject.

[01:04:11:080] Now, we're talking a lot about how to go about doing that in a fair way

[01:04:15:090] when it's impractical to get that consent most of the time.

[01:04:19:100] But, you should understand that that's something that privacy people think about.

[01:04:23:130] And then it, of course, talks about all those other— excuse me— lifecycle things— security, privacy,

[01:04:27:160] and even destroying the data when you're no longer

[01:04:31:170] in need of it. And that also, may be kind of anathema to people who are doing research.

[01:04:35:180] "Well, we're gonna keep it around, 'cause we might think of something later to do with it."

[01:04:39:190] So, the tool kit talks about a variety of these things.

[01:04:43:210] And... um... so... The connection here is the engagement

[01:04:47:240] you know— smoothing over the— the— the— um...

[01:04:51:280] challenges. You can... You can— You can—

[01:04:55:290] certainly do a lot of that by having initially engaged the community

[01:05:00:010] even the subjects. So, if you're doing a study

[01:05:04:030] in a particular neighborhood or a county, trying to connect with

[01:05:08:060] the community there, and figuring out "What can they get?" What are they going to get

[01:05:12:060] out of this work that you're doing. Get them behind it. Often we find,

[01:05:16:080] that some work that states and localities are doing is because

[01:05:20:090] some group of citizens wanted

[01:05:24:110] the data to be able to show something that they're trying to prove—

[01:05:28:140] that they're trying to go to their city council or state government with.

[01:05:32:150] And, they want a study done, so they can find evidence for whatever it is they're trying to promote.

[01:05:36:160] And, if you find those pockets, you may be able to tie in

[01:05:40:170] with those folks. And having that buy-in from the community, from the participants,

[01:05:44:200] is going to smooth over a lot of the kinds of challenges that we're talking about here.

[01:05:48:220] >>MOLLY IRWIN: So, I want to turn to one question

[01:05:52:240] that I think ties in some of the things that you just said, Maya, and some things that have come up

[01:05:56:250] even earlier this morning, and that is about

[01:06:00:260] creating public use data files. I think one of the things that we

[01:06:04:280] are doing more and more in federal government and other folks are doing,

[01:06:09:000] and that's having sometimes, even a task in research contracts

[01:06:13:030] of creating a public use data file at the end of contract.

[01:06:17:040] So, again, we can continue to use and learn from data

[01:06:21:050] that we have. And I wonder if— if— you have

[01:06:25:060] thoughts on what the challenges and solutions are

[01:06:29:080] when thinking about whether and how to create

[01:06:33:110] a public data file, so that we can benefit from

[01:06:37:120] the data that we've worked so hard to get in future projects.

[01:06:41:130] And, I'll start with you, Beth. >>BETH GREEN: Yes, I can comment on this.

[01:06:45:150] So for the first phase of our study

[01:06:49:180] we didn't realize that part of the fine print of our contract with

[01:06:53:210] the CDC was to create a public use data set, until the project was nearing completion.

[01:06:57:220] and so, we had to go back to every single state and amend our

[01:07:01:220] data sharing agreements, which actually was problematic

[01:07:05:240] in a number of cases. They were very reluctant to do this.

[01:07:09:260] And, basically— I mean this is kind of the bad news, in terms of that

[01:07:13:290] part of what the federal government is trying to do. We had to so aggregate

[01:07:18:010] the data for the public use data set, that I really felt like

[01:07:22:020] it's not really useful. Like why would anyone ever want to use this data?

[01:07:26:040] It's not useful. But we did it, because it was in our contract.

[01:07:30:060] And the state agency— a couple of the state agencies

[01:07:34:090] required us to send them the data set— their part of the data set to review.

[01:07:38:100] prior to releasing it as a public use data set.

[01:07:42:120] So, it was a sticking point, and it took a lot of resources to go back and do those [laughs]

[01:07:46:140] amendments. And, I think, ultimately was not that useful.

[01:07:50:160] >>CHARLES MICHALOPOULOS: Um...

[01:07:54:200] Yeah, we were in a better situation, because we knew we had to create a public use file before—

[01:07:58:200] [laughs] beforehand. I talked about some of the concerns that states had

[01:08:02:210] and I think we've generally been able to overcome those.

[01:08:06:230] I think the only issue remaining is that often— OK, we don't have— we haven't

[01:08:10:240] Um... finalized the specifications of the public use file

[01:08:14:270] we have an agreement of principle with each state agency that what we will do will be OK.

[01:08:19:000] But, um... we'll have to make sure that it does follow their regulations in the end.

[01:08:23:010] But I think the key for a lot of them was creating what we're calling a restricted access file.

[01:08:27:030] That will require people signing forms

[01:08:31:050] And really restricting who has access to it

[01:08:35:080] and what they intend to do with it. So it's not really a public use file.

[01:08:39:100] >>BETH GREEN: And, just to add, an example, I mean, I think I went into

[01:08:43:110] administrative data work ten years ago or more, thinking the

[01:08:47:120] de-identified meant take off names and birth dates

[01:08:51:150] social security number, which we never had. And now, to de-identify these data sets

[01:08:55:190] to the level that the states wanted, we had to strip every single date

[01:08:59:190] we had to not have information about individual reports,

[01:09:03:220] even though it was coded. so it's a one instead of a zero

[01:09:07:240] instead of a 99. So, really it's getting down to the level of data

[01:09:11:250] that, like I said, not that useful. Especially, when you get rid of dates.

[01:09:15:280] >>MOLLY IRWIN: So we're nearing

[01:09:19:290] the time that we're going to turn that over to our discussant, but before doing that

[01:09:24:010] I want to give everyone, sort of, an opportunity for concluding remarks.

[01:09:28:020] And, really to focus on what um...

[01:09:32:050] What advice you would give to a new researcher who's

[01:09:36:080] never done this before? What stumbling blocks would you advise

[01:09:40:080] him or her about? And then to talk

[01:09:44:090] maybe a sentence or two about if there was— you know— sort of—

[01:09:48:110] one change that you could make to make it easier to access

[01:09:52:130] administrative data, what would it be?

[01:09:56:160] Let's start with Chuck this time and come this way. >>CHARLES MICHALOPOULOS: Ok, um..

[01:10:00:180] Having to do with researchers, I'd say we already touched upon this. The key thing is having

[01:10:04:190] a relationship with somebody at the state, and having a champion in the state. Where we've done

[01:10:08:210] studies that have been— where the research has been designed in partnership with

[01:10:12:230] the state, anything is possible in terms of getting access to administrative data.

[01:10:16:260] In studies like the home visiting one, where the research had been driven more at the federal level,

[01:10:20:260] it's required a lot more time in negotiations. I think that's the key thing.

[01:10:24:270] I think at the federal level, one big thing is that

[01:10:28:290] we've heard a number of places where perhaps federal agencies could

[01:10:33:010] assemble information that could be available to researchers, so just who should you be contacting at state agencies

[01:10:37:040] to try to do this? What are the regulations

[01:10:41:070] across different states and state agencies?

[01:10:45:090] Uh... uh... And, federal agencies may be in a better position to also build

[01:10:49:100] a relationship over time that we've talked about, rather than researchers who might come in

[01:10:53:120] and need the data from one study, but may not— you know— how where we've collected

[01:10:57:150] unemployment insurance data in California for thirty years, and we have no problem doing that

[01:11:01:160] but, if it's an agency we haven't worked with then there may be

[01:11:05:170] maybe more of an issue. One other—

[01:11:09:190] I know I was only supposed to throw out one, but one other idea, just, a colleague had mentioned,

[01:11:13:210] is wondering whether the federal government could also work with state

[01:11:17:240] attorney generals offices to try to get them to issue cross-agency

[01:11:21:240] um... information about what state agencies are allowed to do so you don't have to

[01:11:25:250] negotiate with each state agency. And then also whether

[01:11:29:260] the federal government could provide researchers with some sort of certification that they could provide

[01:11:33:280] to states saying, "We're nice guys, we'll protect your data." [laughter]

[01:11:38:000] and all that, and maybe that would smooth things over.

[01:11:42:040] >>JENNIFER NOYES: So, I have basically the exact same comments

[01:11:46:050] My big piece of advice, again, comes back to this relationship building.

[01:11:50:060] and when we were first starting out, we would have conversations

[01:11:54:080] with state agencies about— we had a research agenda, so there was "yours, mine, and ours."

[01:11:58:100] There was a research agenda we had. There were questions they were interested in. And then over time it became

[01:12:02:140] much more of an "ours" sort of research agenda. And, it is getting the buy-in

[01:12:06:140] you know, I believe, much as you had said. My wish list

[01:12:10:150] also had to do with the federal level, and it was about

[01:12:14:180] something I touched on this morning, in terms of clearer guidelines on the federal level

[01:12:18:200] about permissions. One of the things that we build into

[01:12:22:240] every single agreement with a researcher to use the data

[01:12:26:250] that we've integrated, is that they have to be able to articulate how it will inform

[01:12:30:260] and improve the program or policy that gave them their data

[01:12:34:280] to get around that, you know— it's collected for a different purpose, but, essentially

[01:12:39:010] yes, it was collected to run the program, but if we can do research that will

[01:12:43:020] help improve the program, then again it's the your, mine, and ours

[01:12:47:030] sort of perspective. But, one of the new challenges is when we go to work with new agencies

[01:12:51:040] or new data sets is to really embed that concept that what we're actually

[01:12:55:050] trying to do is not just a one way street, it's a two way street, and if there was a way to—

[01:12:59:070] I don't know— if it's a certification— you know, someone saying we're good guys and we're here to help,

[01:13:03:100] and all of that, I don't know. But definitely because things are all over the map,

[01:13:07:100] and you're renegotiating, over and over again, what you thought was a concept you had figured out.

[01:13:11:110] If there was just like one go-to place— like a tool kit like you were talking about

[01:13:15:130] or something like that, that would be really nice.

[01:13:19:160] >>MAYA BERNSTEIN: What would I tell a new researcher? Um...

[01:13:23:190] I guess, I would try to say: bake the privacy in. [laughs]

[01:13:27:200] Start at— I maybe stole my own thunder from the beginning, right?

[01:13:31:210] Think about the entire life of the project that you're doing at the beginning, and try to,

[01:13:35:220] at each step, know what the privacy rules are at those steps,

[01:13:39:250] and plan for that in advance.

[01:13:43:280] I'm not sure you really want me to answer. [laughs]

[01:13:47:280] How can I make access to data easier? But, I will. [laughs]

[01:13:51:280] And I guess I would say that— try to be— try to anticipate And this is part of—

[01:13:55:300] Well, I have two things, really. Try to anticipate what your partner

[01:14:00:020] in the project is dealing with. You mentioned

[01:14:04:040] legal matters, and policy matters, but also budget issues.

[01:14:08:080] And, try to anticipate those, and have thought about them,

[01:14:12:080] before you approach those people. And then, I think,

[01:14:16:120] for me, the one thing that— It comes up— It came up

[01:14:20:140] less often here, but I hear often— I wish I could change the language about how we talk about this.

[01:14:24:160] You know, I wrote down some of the words that I heard here.

[01:14:28:190] Barriers. Hoops. Impediments. Road blocks. How can we get around the problem, right?

[01:14:32:220] Which— which is just a way of speaking, but for me

[01:14:36:230] the privacy rules, as I said before, are the way you get the data.

[01:14:40:260] Because, if people are not confident, as I said before, you're not going to

[01:14:44:290] get the data at all. From a health point of view, if people are not, for example,

[01:14:49:000] comfortable talking to their physicians, and candid with them,

[01:14:53:010] you're— you know— We have a public health problem on our hands, right?

[01:14:57:030] If people don't tell their doctors about their STDs or their HIV status

[01:15:01:050] there are actually real consequences

[01:15:05:090] to that. So, I wish we could change the language around this, and just say

[01:15:09:110] this— you know, I don't like the idea of— we talk about

[01:15:13:130] privacy balancing things. I really don't like that metaphor, because it implies that

[01:15:17:140] if we have more privacy; we have less of something else. Or we have more data access; we have less privacy.

[01:15:21:170] I much prefer the metaphor that a rising tide lifts all boats.

[01:15:25:200] We can get both. We can do things in a privacy protected way

[01:15:29:210] and get the research done. And that's how we should— That's how I wish

[01:15:33:220] we would think about the problem more.

[01:15:37:240] >>BETH GREEN: I don't have much to add, [laughs] to what has already been said.

[01:15:41:270] I would echo the— anything that could make the process of access

[01:15:46:020] more consistent, either across agencies within a state,

[01:15:50:030] or across states, would be such a huge advantage.

[01:15:54:050] And, by consistent, I mean: Do you have to go through another IRB?

[01:15:58:070] Do you have to— What kind of security do you have to have on your hardware,

[01:16:02:100] software? What kind of language needs to be in there? Anything like that, I think

[01:16:06:110] would be helpful. In terms of what I would tell a new researcher,

[01:16:10:130] I guess I would say, and this has to do with relationships, see if you can find

[01:16:14:140] a researcher who's already built that relationship and tag on.

[01:16:18:160] And work with them, because they probably know who to talk to.

[01:16:22:220] One of the things that we started doing in Oregon— although, I'll say that we have had

[01:16:26:230] limited success, but— We got a group of researchers together, all of whom were at various

[01:16:30:240] processes for various projects, asking our state agency—

[01:16:34:260] Child Welfare Agency, for data. and we thought,

[01:16:38:290] maybe we could coordinate our requests better, so that the state

[01:16:43:020] doesn't feel like they are being bombarded with a million different requests

[01:16:47:030] for a million different projects. at least in terms of figuring out: What is a core set of data elements

[01:16:51:040] that we're all interested in and can all agree on? And, I actually thought that had some

[01:16:55:060] promise, in terms of, let's work together

[01:16:59:090] since there's now quite a few researchers in the state who would like to have access

[01:17:03:120] to this data. So, I'm hoping we can move that along. But, that might be another

[01:17:07:130] thing to help with access. >>MOLLY IRWIN: Great. Let's keep going

[01:17:11:150] right down the line to Kelly, who's now going to serve as our discussant.

[01:17:15:180] >>KELLY MAXWELL: Thank you. This has been a great discussion, and I'm very excited to be here

[01:17:19:210] for all today and tomorrow morning. So, my job was just

[01:17:23:210] to highlight a few of the take-aways that I had from this conversation.

[01:17:27:220] And, so, I will admit that I had the benefit of a planning call,

[01:17:31:230] that was also a very intense discussion, as well as the comments that

[01:17:35:250] were made today. So, I think the first take-away point I have is that

[01:17:39:280] access is like an onion. There are many layers to it,

[01:17:44:020] and I hope it doesn't make us cry when we're working on it. [laughter]

[01:17:48:020] The first layer I want to mention is availability. And, I think, sometimes

[01:17:52:040] we don't think about that. But, access is really availability. Do the data exist?

[01:17:56:070] Does the program or agency gather the the data,

[01:18:00:100] or do they gather something that is kind of similar

[01:18:04:110] to what you might be able to use, but not quite? So, availability is the first

[01:18:08:120] layer for me. The second layer, is physically getting the data–

[01:18:12:140] And I think that's often, what we think of first when we think about access.

[01:18:16:160] And that's the things like establishing data sharing agreements and many other steps–

[01:18:20:200] actually getting the portion of the data set that you need to answer your question.

[01:18:24:220] The third layer, though, is often times our questions

[01:18:28:230] require not just data from one data set, but actually merging

[01:18:32:250] data from multiple places in order to have all of the data we need

[01:18:36:270] to answer the question. So, I think the third layer is really

[01:18:41:000] the issue of merging the data, in order to truly have access to the information you need.

[01:18:45:020] And I'm looking forward to the session that, I think, is at the end of today,

[01:18:49:020] that's going to talk a little bit about challenges and opportunities with linking.

[01:18:53:040] And then, finally, the last layer, for me– The fourth layer is understanding what the data

[01:18:57:060] mean and don't mean. And, that is a critical issue

[01:19:01:100] that Maria's example this morning was fabulous about.

[01:19:05:110] She was ready to go down that road about understanding why

[01:19:09:120] people were moving. And, if she hadn't have gone and talked to that person,

[01:19:13:140] she wouldn't have really understood all the issues that surround why that

[01:19:17:160] particular code was coming up so important. So, I think understanding

[01:19:21:190] the meaning, for me, is about understanding

[01:19:25:200] the quality of the data that are collected, the quality and how

[01:19:29:210] the data are entered into the system, how the variables were defined,

[01:19:33:230] as well as the policy context. And all of those layers are really important

[01:19:37:250] when we think about access. The second thing is

[01:19:41:280] relationships are critical. As everybody has said here, and as we know

[01:19:45:280] throughout lots of aspects of our lives. And, I think it is really

[01:19:49:290] critical to develop trusting, mutually beneficial

[01:19:54:000] relationships among all the stakeholders who are necessary

[01:19:58:020] in order to use administrative data to answer important questions.

[01:20:02:040] Jennifer talked a lot about taking the time when you're working with—

[01:20:06:080] within a single state or a single agency.

[01:20:10:090] Many things you can do to develop a long term relationship. Chuck has a different

[01:20:14:100] challenge, and I think, in my head, it's kind of like speed dating.

[01:20:18:130] He had to develop relationships with a lot of agencies in a lot of states.

[01:20:22:150] So, I know I, for one, am looking forward to the informal networking session tonight,

[01:20:26:190] because I am going to ask him a little bit more about that, because I think he has lessons

[01:20:30:200] learned for us, who are doing— who are interested in doing

[01:20:34:220] multiple state work with administrative data.

[01:20:38:240] My third take-away is related to what

[01:20:42:270] Maya said at the end here, and I want to use the phrasing she used,

[01:20:46:270] on the planning call, which is: it's both/and, not either/or.

[01:20:50:290] That really we shouldn't think about privacy and confidentiality

[01:20:55:010] as the bad guy. That really we need to think about,

[01:20:59:030] in order to do our research well, we have to think about

[01:21:03:060] both confidentiality and privacy and access issues.

[01:21:07:080] The fourth take-away is that we need

[01:21:11:090] to build capacity among all of us— researchers,

[01:21:15:110] administrative organizations that house the data, and funders.

[01:21:19:140] And capacity was the first thing that Beth mentioned today when asked about challenges.

[01:21:23:170] The amount of data that we have available is growing

[01:21:27:180] as is the technology that we have to use it, and the analytic techniques that we have.

[01:21:31:200] As well as the laws and policies that govern the use of data

[01:21:35:210] is growing and likely to continue to grow and evolve. So, I think it's a big challenge

[01:21:39:230] for all of us to just keep up with all of this.

[01:21:43:260] And, I think we need to think about building capacity at multiple levels of

[01:21:47:280] this system. So thinking about capacity at the local program level in the community

[01:21:51:290] at the county level, at the regional level possibly

[01:21:56:010] at the state level, and federal level. I also think we need to think about

[01:22:00:030] multiple aspects of capacity. So, I'm sure there are

[01:22:04:060] millions of aspects of capacity, but I'm going to mention a few here. And, the first one I want to

[01:22:08:070] start with is data systems. Do the programs and organizations

[01:22:12:080] have the data systems to support their storage and use of

[01:22:16:090] administrative data? My guess is that most programs and organizations

[01:22:20:110] are really limited with what they can do because of their data systems

[01:22:24:120] The second, um...

[01:22:28:160] thing I want to mention is expertise. And when I think about expertise,

[01:22:32:170] I typically think about the expertise in the organizations

[01:22:36:180] that have the administrative data. That they either have to have staff in house,

[01:22:40:200] that understand what the data mean and how to use it

[01:22:44:230] and analyze it, or they need to partner with somebody outside of their organization

[01:22:48:270] who has that expertise, who can help them utilize the data

[01:22:52:280] to answer their policy question. And that brings me to

[01:22:57:010] the fourth— third aspect I want to mention, which is related to that

[01:23:01:030] but more broadly, about culture. I think part of capacity is

[01:23:05:060] this broad organizational culture that values data

[01:23:09:070] and values the use of data in making decisions. So, that's another aspect.

[01:23:13:080] And then, finally, the last thing I want to mention is knowledge.

[01:23:17:090] And, of course, we all have to stay up to date in particular areas

[01:23:21:110] of our expertise that we work in. And, I think we all

[01:23:25:140] have to stay knowledgeable about the laws and policies

[01:23:29:160] that govern the data. So, we can turn to experts

[01:23:33:170] and maybe our new best friends, like Maya and the friendly lawyer John,

[01:23:37:190] [laughter] who can provide advice about that. And, there are also

[01:23:41:210] a growing number of resource out there to help us and our partners

[01:23:45:240] understand these issues. So, I just want to mention one of those today.

[01:23:49:260] Research Connections promotes high-quality

[01:23:53:270] early care and education research, and the use of that research.

[01:23:57:290] So, even though they are focused primarily on early care and education,

[01:24:02:020] I think some of the resources that they have really apply to broader areas.

[01:24:06:040] And, I want to talk today about one of the pages that

[01:24:10:070] they have on their website, that has pulled together a variety of resources

[01:24:14:080] about using administrative data. So, I think part of the challenge is:

[01:24:18:100] I might know about a resource that's over here, but I might not know about

[01:24:22:140] a resource that's over there. Or our state partners might not even really know: where do I turn

[01:24:26:180] if I want to learn a little bit more about how to set up a data sharing agreement?

[01:24:30:180] So, Research Connections has tried to pull a range of resources together

[01:24:34:200] and organizing them. So, they have brief descriptions and links

[01:24:38:220] of things like the ACF Confidentiality Tool Kit. And, they also have

[01:24:42:250] links to other websites, like the Privacy Technical Assistance Center, that you could–

[01:24:46:280] kind of while you're there thinking about all of these issues, click and go to other

[01:24:50:290] websites that are related to using administrative data.

[01:24:55:020] And then, finally, I want to end with, again,

[01:24:59:040] a point that– >>JENNIFER NOYES: Can you tell us– before you– can you tell us more about

[01:25:03:070] who Research Connects are? Who are they? >>KELLY MAXWELL: [talking over] So,

Research Connections is funded by

[01:25:07:080] the Office of Planning, Research, and Evaluation– That's us! [laughter]

[01:25:11:090] And you can go to– I think it's www.researchconnections.com

[01:25:15:100] >>AUDIENCE VOICE: Dot org... >>KELLY MAXWELL: Dot org... Oh, sorry. Thank you.

[01:25:19:120] Um... they have a well– I mean it's primarily focused on

[01:25:23:150] early childhood research, but again, because of this issue of–

[01:25:27:160] there's a growing number of resources out there about working with administrative data,

[01:25:31:160] but they're all in multiple places– We are trying to pull things together

[01:25:35:180] >>JENNIFER NOYES: Thank you. >>KELLY MAXWELL: Yeah, talk to somebody at OPRE

[01:25:39:190] uh, maybe Maura, right? You're the person who

[01:25:43:220] if you had more questions about Research Connections... And, again–

[01:25:47:250] Finally, I want to end with the point that Maya said on the planning call,

[01:25:51:260] which is, to remind us that there's a person behind every data point.

[01:25:55:270] And, I think when we are reviewing spreadsheets or statistical reports,

[01:25:59:290] that can be something that we forget. But, I hope we will remember

[01:26:04:020] during this meeting and beyond, the children, families,

[01:26:08:050] service providers, and others, who are behind the data that we're using. Thank you.

[01:26:12:080] [applause]

[01:26:16:100] [applause]

[01:26:20:130] We have almost fifteen minutes for questions. There are

[01:26:24:160] microphones up front.

[01:26:28:170] So, if folks have questions, please come forward,

[01:26:32:180] and... um.. and... we will keep going.

[01:26:36:200] And, I guess, while people are coming forward I'll give

[01:26:40:220] a minute to see if panelists have questions for each other.

[01:26:44:250]

[01:26:48:250] I think so, yeah...

[01:26:52:260] >>MAYA BERNSTEIN: You might have to get closer to it. [laughter]

[01:26:56:280] Flip the switch. >>MOLLY IRWIN: It's not working?

[01:27:00:290]

[01:27:05:020] >>AUDIENCE MEMBER: I'm good now?

[01:27:09:050] >>MOLLY IRWING: Oh yeah. >>AUDIENCE MEMBER: OK. Thank you. This has all been really interesting

[01:27:13:060] and helpful. I wanted to ask you a little bit more, to talk about data

[01:27:17:070] ownership. And, I heard two things referred. One was

[01:27:21:090] the owner of the data, but also Maya used the term custodian of the data.

[01:27:25:120] And, the fact that the privacy rules might change with the custodian of the data.

[01:27:29:150] So, what's the difference between a data owner and a data custodian?

[01:27:33:160] And, I ask this partly, because as a faculty member

[01:27:37:180] at a public university, I'm advised that when I

[01:27:41:200] get data through a data sharing agreement, that I also sign–

[01:27:45:240] It has a great name. It's a materials transfer agreement.

[01:27:49:260] I think that's what it's called. Basically it says

[01:27:53:260] I don't own the data that you've just given me. Because, if someone files

[01:27:57:280] a Freedom of Information Act and asks me for my data

[01:28:02:000] because I'm at a public university regardless of whether

[01:28:06:030] or not it was given to me with whatever confidentiality agreement

[01:28:10:040] I've signed. So, no one's ever tested that, as far as I know,

[01:28:14:050] but, anyway, I wondered if anyone else had that experience but also, just, this question

[01:28:18:070] about ownership versus custodian. >>MAYA BERNSTEIN: So, maybe I should jump in. Thank you.

[01:28:22:100] Um... Because I'm a lawyer, when I think about the term ownership,

[01:28:26:100] that implies to me all kinds of property rights and other kinds of very complex—

[01:28:30:120] a long, you know, list of property laws that goes back to, you know...

[01:28:34:130] civil law in England, or something. And, I don't usually

[01:28:38:140] think about data that way, because— Well, first of all, if you're at a public university

[01:28:42:160] or here we are at the federal government— Ownership? I mean, who owns the data? The taxpayers own the data.

[01:28:46:180] Right? Really. We're collecting it on behalf of the taxpayers to do work for them.

[01:28:50:220] I did use the word ownership, I realize, earlier, and I meant, you know,

[01:28:54:230] in the colloquial way— The program person owns their data,

[01:28:58:250] because they're responsible for it. But, I try to use the word custodian

[01:29:02:270] Because... um... Let me give you an example. In the HIPPA privacy law,

[01:29:07:000] there's a lot of people who wanted us to say— well, who owns the data?

[01:29:11:040] Well, if you think about it, if you're the patient, it's your data.

[01:29:15:040] It's about you, right? But it's the result of your doctor's

[01:29:19:060] intellectual work. It's, uh.. you know...

[01:29:23:080] You could make an argument that it's owned by the insurance company that paid for those services.

[01:29:27:100] You could make an argument that it's the hospital who provided the facilities.

[01:29:31:100] It doesn't make sense to really think about ownership, because so many different people

[01:29:35:110] have a— so many different parties have an argument that they "own" the data.

[01:29:39:120] So, most privacy law, and I think maybe

[01:29:43:140] the way we can think about it is if you've got the data right now,

[01:29:47:160] and you're the— another word— a steward of that data, right?

[01:29:51:190] So, in the tool kit, we talk about data stewardship. How is the proper way to manage that data

[01:29:55:200] while you're the person responsible for it. And, so, it's true, I—

[01:29:59:220] Ownership is— I find, not very useful as a metaphor.

[01:30:03:230] [several speaking] >>KELLY MAXWELL: Can you talk more

[01:30:07:270] about what you meant by custodian?

[01:30:12:000] >>MAYA BERNSTEIN: I guess I just meant the person who is responsible for that data where it is

[01:30:16:010] at the moment. And that might be— for a federal agency, that's the federal program office.

[01:30:20:020] For a state, that's the state agency.

[01:30:24:040] For a university, that could just be the PI for that particular project.

[01:30:28:070] So, it could be at different levels. But, the custodian

[01:30:32:100] is, I guess, at bottom, who is the person whose neck is on the line if something goes wrong?

[01:30:36:110] Right? [laughs] >>JENNIFER NOYES: The only comment

[01:30:40:140] that I would make, also, at a public university is, we're really very clear with

[01:30:44:160] the data– with the merge data set that we've created, that I guess the right language would be

[01:30:48:190] that at that point in time we're the custodians, but we don't have

[01:30:52:200] ownership– It's the state agency's data.

[01:30:56:210] So, in order for it to be used, we need to go back to the state agency from whom we got it.

[01:31:00:220] Because, we have the whole population across multiple systems, to get permission to use it.

[01:31:04:240] So, we cannot, just release it. Right? So, we're in the same situation

[01:31:08:260] that it sounds like you're in. But, it is interesting

[01:31:12:300] to think about this custodian language, because clearly part of

[01:31:17:000] when I talk about relationships and relationship building over time,

[01:31:21:020] I personally feel like I have an obligation to be very aware of

[01:31:25:040] and professionally cognizant of my responsibilities and relationship to that data

[01:31:29:070] And so, therefore, it just can't be used willy-nilly,

[01:31:33:100] because, I'm the custodian at that point. So, I understand the language that you're using in terms of

[01:31:37:110] my responsibility as– >>MAYA BERNSTEIN: Right, and I think the other thing to say is that

[01:31:41:130] responsibility doesn't necessarily come from the organization who is the custodian, right?

[01:31:45:150] It could come from state or federal law. It could come from your

[01:31:49:180] data use agreement– so that's a contract kind of thing– it could come from your university's policy,

[01:31:53:200] it could come from a lot of different places that you're having to manage, but as the

custodian

[01:31:57:210] you're managing whatever the imposed rules are for that data

[01:32:01:220] at that time. >>BETH GREEN: I would just add,

[01:32:05:250] I can't really speak to the ownership. I think our understanding's— also a public university—

[01:32:09:260] is similar to Jennifer's. That we are the custodians of the data.

[01:32:13:300] That said, I'm not familiar with the process that you talked about

[01:32:18:010] in terms of protecting the data from subpoena. We do do a lot of work with

[01:32:22:030] child welfare agencies, so we do get subpoenaed for data occasionally— a handful of times

[01:32:26:050] in the past five years. And typically with administrative data,

[01:32:30:080] we just say this is the— We have the data that the state has.

[01:32:34:090] So, if you want this data, you already have it, essentially. [laughs]

[01:32:38:100] And so, they've never pushed us on that. >>MAYA BERNSTEIN: Finish your thought, but you reminded me of something else.

[01:32:42:120] >>BETH GREEN: And then, I was just going to say, other kinds of data that we've collected as researchers through interviews and things

[01:32:46:150] things like that, we do have a federal certificate of confidentiality that protects it from subpoena

[01:32:50:180] So, we have used that mechanism. >>MAYA BERNSTEIN: That's what

[01:32:54:200] I was going to— about certificates of confidentiality.

[01:32:58:210] >>MOLLY IRWIN: Alita... [inaudible]

[01:33:02:220] >>MAYA BERNSTEIN: You have to get closer.

[01:33:06:230] >>MOLLY IRWIN: Speak loud.

[01:33:10:240] >>AUDIENCE MEMBER: Where's the— Ooo! >>MAYA BERNSTEIN: You jsut have to get

close to it.

[01:33:14:260] >>AUDIENCE MEMBER: So, um I work with a number of folks– Maria Woolverton is leading this project,

[01:33:18:290] where we're working specifically on creating a

[01:33:23:040] nationally representative of American Indian and Alaskan Native

[01:33:27:040] Head Start centers, and I did notice

[01:33:31:060] that no one– I did not hear when I was in the room, anybody talk about issues around

[01:33:35:080] working with sovereign nations, working with tribal communities, and one thing I

[01:33:39:120] was just so happy to hear you all talk about was this issue around

[01:33:43:130] what the data mean and don't mean, and really thinking about

[01:33:47:140] interpretation with the stakeholders, and, sort of, the moral responsibility to make sure that data

[01:33:51:160] are not interpreted without input from those stakeholders. So, I was wondering

[01:33:55:180] if.. um... I guess, I see Beth nodding. I was wondering if

[01:33:59:210] any of you would be willing to share if you have done some work with

[01:34:03:210] sovereign nations around these issues. I think it's important to make sure we talk about today.

[01:34:07:230] >>BETH GREEN: I haven't personally, but have colleagues who do. And I know that

[01:34:11:240] the– my understanding, at least, is that the sovereign nations

[01:34:15:260] own their own data. And, that that

[01:34:19:270] is a totally separate process for– even if they're part of a federal grant.

[01:34:24:000] And there are federal officers here who can speak to this more than I. That they still own their data,

[01:34:28:010] so there's a whole separate layer of process that has to go on, in terms of working with

[01:34:32:020] American Indian nations and tribes. >>MAYA BERNSTEIN: Can I add to that?

[01:34:36:050] So, during the process of working on the tool kit that I mentioned to you,

[01:34:40:070] The National Committee on Vital and Health Statistics, we did hear from

[01:34:44:110] representatives of the tribes. And, a couple

[01:34:48:120] very interesting things. Are there people in the room who are familiar with the Havasupai case?

[01:34:52:140] A few of you. So I'll tell you briefly. The Havasupai are

[01:34:56:140] the tribe that lives in the basin of the Grand Canyon.

[01:35:00:170] And, there was a famous research study at Arizona State

[01:35:04:210] that was looking into and taking samples

[01:35:08:220] of the Havasupai people, with their permission, on diabetes

[01:35:12:240] which is rampant in that community. And the data were later

[01:35:16:250] repurposed and used to do a study of schizophrenia and also used later

[01:35:20:280] to do a study on migration patterns.

[01:35:24:290] that published a result that differed from

[01:35:29:000] the Havasupai's own story about their origins.

[01:35:33:020] And, the tribe was really unhappy. Sued Arizona State, and they paid out

[01:35:37:030] a \$750,000 penalty for having done that without the proper authority

[01:35:41:060] So, two things from that— some of the things that I've already said about...

[01:35:45:100] Is, as you said, to involve the

[01:35:49:100] the community and the people that you're doing with and remember that there's

[01:35:53:110] people underneath this. But, something that we didn't mention—

[01:35:57:130] mentioned in passing, which is that—

[01:36:01:160] it's not exactly privacy, but there is kind of idea of the privacy of a group of people.

[01:36:05:200] Right? Not individual privacy, which I usually think about, but, is there something

[01:36:09:210] to saying something about a particular group that stigmatizes

[01:36:13:220] that group. And, so, how do you think about reporting out

[01:36:17:250] there's a high incidence of schizophrenia

[01:36:21:280] in this population, or a high incidence of alcohol in this population?

[01:36:25:290] Does that stigmatize that population? How— What happens to that population

[01:36:30:000] as a result? In their community or with other people that they deal with,

[01:36:34:010] if it's known? Does that advantage or disadvantage that group?

[01:36:38:030] And, so, we have to think about the reporting out of that kind of result as well.

[01:36:42:060] Which goes to some of the things that— in particular, small communities like tribes

[01:36:46:080] if they're the small ones— do care about quite a lot. And the same issue

[01:36:50:090] I guess we mentioned when we talked about the state known as the

[01:36:54:100] you know, the last in whatever it is we're measuring

[01:36:58:130] for the same kinds of reasons.

[01:37:02:160] >>AUDIENCE MEMBER: Hi. I'm Nicole Deterding. I'm a post-doc at OPRE.

[01:37:06:200] And so, I'm coming, sort of from just finishing a Phd, and being in academia more,

[01:37:10:200] and I think today, people here have been, for very good reason,

[01:37:14:220] both practical and ethically on the side of building

[01:37:18:240] mutually beneficial relationships with the people who own the data. But there is, sort of, out there

[01:37:22:270] in the world, another sort of

[01:37:26:280] thing that I think is in tension with that, which is we need freer access to

[01:37:30:290] administrative data. Um... I was at a

[01:37:35:010] hill briefing the other day, and a very prominent academic sort of told the group

[01:37:39:040] "Go home and tell your bosses we need more access to administrative data."

[01:37:43:070] So, what would you say very directly to the people who are pushing for that,

[01:37:47:090] and in a world in which data owner's time is limited

[01:37:51:100] um... How do you balance

[01:37:55:110] the time that it takes to build these really strong relationships, and maybe freer access to the data?

[01:37:59:140]

[01:38:03:170] [muttering] >>MAYA BERNSTEIN: So, I have

[01:38:07:180] a little note here, actually, that's sort of on that topic, which is that

[01:38:11:200] certainly the administration is promoting

[01:38:15:220] public access to data. We have data.gov, and other ways

[01:38:19:250] of getting data out to the public. And for the purpose—

[01:38:23:280] Basically, for whatever purpose people might make of it. The latest cool app, or whatever it is.

[01:38:27:280] But, I think that's in tension

[01:38:31:290] with the idea that

[01:38:36:010] when you make inferences from data— out there unstructured and out—

[01:38:40:060] well maybe not unstructured, but without information

[01:38:44:090] about what that data is and what that data was collected for

[01:38:48:100] and what the limitations of that data are, without some kind of metadata about what does this data mean?

[01:38:52:110] Which we've talked about a little bit.

[01:38:56:120] People who've freely accessed the data because they're making the latest cool app,

[01:39:00:140] or trying to make some cool study, without a sophisticated

[01:39:04:180] knowledge of what the data is, you can make inferences of data

[01:39:08:190] that were collected for another purpose, but that's dangerous,

[01:39:12:210] because you may find that those inferences

[01:39:16:240] are just wrong. That they're just not— You can find correlations

[01:39:20:260] that have nothing to do with causation, and so forth. We're familiar with that problem.

[01:39:24:300] And so there is, I think, a tension between those two things. We do think about that a lot

[01:39:29:010] at HHS in the federal government when we talk about making public use data files.

[01:39:33:030] And we talk about disseminating data. We're thinking about how to better create that

[01:39:37:050] that metadata so that there's a way to track back

[01:39:41:080] what was that data collected for and what did it mean at the time it was collected, because when you use it for another purpose

[01:39:45:080] you are going away from what the meaning of that data is.

[01:39:49:090] And, that is not necessarily a good thing. >>JENNIFER NOYES: And as a follow up

[01:39:53:100] to that, I think one of the ongoing things that we try to manage day to day,

[01:39:57:120] is this balance, and in so much as, as I was saying before,

[01:40:01:130] as a researcher, you're just interested in what I can the one-off

[01:40:05:160] like you just want to get the data to accomplish what you want to accomplish

[01:40:09:160] and it can— I don't know if the right word is to say ruin for the good of the order

[01:40:13:170] but, you have to remember there's a bunch of us that want to have access to that

information

[01:40:17:190] and be able to use it in a responsible manner to inform policy and practice, which is what I'm all about

[01:40:21:210] And, you can damage relationships that somebody else has

[01:40:25:220] built over time by inappropriate or what you think is appropriate

[01:40:29:230] but sort of cavalier use of the data.

[01:40:33:230] So, I do think there's this really strong tension and I'm all for

[01:40:37:260] I'm a "give me everything you have" type of a person and then back off of it, you know?

[01:40:41:280] Um... as far as the data goes. But, I'm also super protective of understanding.

[01:40:46:010] I had it in my notes as well, the concept of: remember these are people on the other end of the data.

[01:40:50:050] This isn't just about fun and games with research.

[01:40:54:060] And then we have to remember that, and then we have to remember that this again should be all about

[01:40:58:090] improving programs and policies in practice, and not getting just

[01:41:02:120] the next journal article. Which is a conversation that I frequently have.

[01:41:06:150] It's about the journal article. It's about what can be done to improve policy and practice for the people

[01:41:10:150] whose data that we're using. Even if it's not in the direct instance, like when you write your IRB protocol—

[01:41:14:170] How is this going to benefit the individuals that you're studying? And, really it's not.

[01:41:18:190] but over time people like them it could be benefiting. And so, again,

[01:41:22:220] it is a really strong balancing act. And, just this free, unfettered—

[01:41:26:250] without this basic understanding, to me is very problematic and actually threatens

[01:41:30:260] our work going forward; it doesn't enhance it. >>MAYA BERNSTEIN: We did, in the

hearings

[01:41:34:280] that we did with the national committee that I mentioned, hear a lot about this

[01:41:39:000] concept of helicoptering researchers

[01:41:43:030] who come in, collect the data, publish their paper, put it on their CV, get their tenure,

[01:41:47:060] and we never see them again. And, I think, perhaps we need

[01:41:51:070] a new way of thinking about this. Maybe we need to think about the ethical obligation of

[01:41:55:080] using the data and feeding that back to the community

[01:41:59:100] or the people who are the subjects of that data.

[01:42:03:140] >>MOLLY IRWIN: Great. On that note, it brings us, I think, about exactly to noon.

[01:42:07:170] So, first join me in thanking

[01:42:11:180] our fabulous panelists. [applause]

[01:42:15:200] [applause]

[01:42:19:220] Lunch follows. It's, uh... We should be back by 1:30.

[01:42:23:250] There are a number of places around here. Is there a list somewhere?

[01:42:27:260] There are a number of places around here, and grab the arm of

[01:42:31:270] somebody who works close by, and they can help point you.

[01:42:35:290] We should be back by 1:30. Everybody should take their stuff when they go.

[01:42:40:010] And we'll see you at the next session at 1:30.