

>> Hi there, and thank you for joining us for today's digital impact virtual roundtable. Today we're looking at data labs, improving access to government data. I'm Lucy Bernholz, Director of the Digital Civil Society Lab at the Stanford Center on Philanthropy and Civil Society. These virtual roundtables are a community-generated set of conversations about issues related to digital data in civil society. The conversations are part of a larger constellation of activities that we host at the Digital Civil Society Lab and through our Digital Impact Initiative. The multiple communities that we engage with include our international network of digital impact grants awarded [Inaudible] and applicants. And the many fellow scholars, practitioners and leaders we've met and learned from on our digital impact world tour. Which has taken us across five continents over the last year. And our next step is in Medellin coming up later this month. So a warm welcome to everyone from those spaces and elsewhere. And we invite you to learn more about all of these initiatives through the digitalimpact.org site and the Digital Civil Society Lab. Today we're going to discuss an emerging data-sharing model that helps dissolve barriers to collaboration while safeguarding sensitive data. The Data Labs Project, led by New Philanthropy Capital and the Governance Lab aims to make government administrative data more useful. So that nonprofits can better understand the impact of their services on beneficiaries. As we all know, governments collect and store massive amounts of data about a wide range of issues, from public health data to crime stats, to student test scores. The information is often referred to as administrative data. And these data are a product of government basically doing its job. And they can be a potential input to inform public policy. Nonprofits often need better information about government actions in order to plan, provide, and evaluate their services. Government data can be useful to nonprofits, but it's often not easily available within government or outside of it. Regulations that are meant to keep government in check can limit how data are shared and who has access. Privacy laws that are meant to protect citizens from government overreach play an important part in this. And nonprofits that can benefit from seeing their work in light of broader public data sets may not have the analytic skills or capacity to actually make sense of the massive quantities of raw data. So it's worth imagining better ways to take advantage of the data we have and the nonprofit wisdom on the ground. How can citizens trust government to do right by their data while also taking advantage of the opportunities to use data to improve services in communities? In this conversation, we'll be hearing a couple of examples of ways to strike this balance. Before we get into the discussion, let me go over a few housekeeping details. For everybody but the panelists, your microphones will be muted for the length of the discussion. We do want to hear from you and we're delighted that some of you have already sent in questions ahead of time. In order to participate, please use the question function on your control panel to chime in, so that your questions starting now or at any time throughout the conversation. And I will pass them on to the panelists and everyone on the call. We'll try to get to as many of your questions as we can in our limited time today. Don't hesitate to send them in immediately. Your questions are critical to this conversation. We will be recording and sharing out this dis-

cussion on the digital impact podcast, which is available on iTunes and on the website, digitalimpact.org. And there you can also find a host of other resources for using digital resources safely, ethically, and effectively in the social sector. We've also got more virtual roundtables planned over the next coming months. So stay connected with us on social media, subscribe to the mailing list for updates, and we'll keep you informed of what's coming next. Remember, these conversations are driven by and for the community. So if you have topics or questions that you'd like to see us explore, email us at hello@digitalimpact.org, and we'd love to host a conversation on the topics that you want to talk about. So let's get started. Let me first introduce our panel. We're joined by Tracey Gyateng, who is the Data Labs Project Manager at New Philanthropy Capital. Tracey, say hi to everybody. Rod Clark is the Chief Executive, Prisoners' Education Trust. Hi, Rod.

>> Hi there.

>> And Justine Hastings, who is the Director of the Rhode Island Innovation Policy Lab, or RI IPL. And thank you very much to each of you for joining us today. Let me try to frame up the conversation a little bit. The primary goal here at Digital Impact is to advance the safe, ethical, and effective use of digital resources for the benefit of civil society. Today we're talking about this emerging data-sharing model which is trying to shift or dissolve or ease the barriers to using government data and what that can mean for the social sector. This is not a static topic. There's new legislation in the US that's been proposed that would establish both a more secure and a transparent data system to help federal agencies assess the effectiveness of their own programs. In the EU, a new law, the GDPR which is set to redefine how corporate data collectors protect citizens' personal data is coming on board in May. And despite open data initiatives gaining widespread support, concerns over privacy and the ethical use of personal data have only increased. Tracey Gyateng and Beth Simone Noveck recently wrote in a blog post for the Stanford Social Innovation Review, "We need the efficient production and use of rigorous evidence to become a routine part of government operations and public policy makings." So the question is, are data labs resilient enough to handle the rapidly changing regulatory environment? As well as, how do they work? And how do you get to be part of them? Today's panel is here to tell us a little bit about where the data model came from, how it's working, how it isn't, and where it's headed. So let's get started. Tracey, I'm going to start with you. New Philanthropy Capital started the Data Lab Project to help ensure that more access to government data actually helps nonprofits measure their impact. The Justice Data Lab was launched by the UK Ministry of Justice five years ago. Tell us a little bit about how the model works, what you're working on now, and how you think about privacy at the Justice Data Lab.

>> The Justice Data Lab supports organizations who are working with offenders or ex-offenders. And helps them to kind of understand whether the interventions or services that they provide them actually support people to stop reoffending

or reduce the frequency of their offending. And so essentially how it works is that these organizations will have personal data of the people that they have collected from, their names and date of birth. And they will securely send that data to the Ministry of Justice through an encrypted network. The Ministry of Justice will then link that data to the police national computer database. And then they will calculate for that group of people that the organization has worked with various reoffending metrics. So, have the group reoffended – yes, no? The frequency of offending. Sometimes the seriousness of offending. And the length of time it takes to reoffending. Because in the criminal justice world these are important metrics; it's not just about whether you stop reoffending, yes, no. But over time you might actually reduce your offending. And so the charity then gets information which tells them, what's the outcomes of the people that they've worked with? So I'll give an example that Rod will be able to speak much clearly on, but just as a make-believe example, we could just say I've submitted data on a hundred people and 50% of them went on to reoffend. So now I know what happened to them. But actually what's really interesting to know is, well, what's the impact? What's the difference that my organization has actually made? And with that, you need a comparison group. And so the Justice Data Lab, what the Ministry of Justice is able to do is to match a cohort of people that share similar characteristics to people that the organization has worked with, and calculate the reoffending rate for this group. So in my make-believe example, let's just say 80% of people that shared similar characteristics to the people that I have worked with have reoffended. And so I'm able to compare the difference and see though actually I was 50%, the comparison was at 80%, and there's a 30% difference. And we can kind of attribute that in a way to the service that has been provided. And what's nice about this project, the model, is that actually privacy is actually encapsulated within the model. So in the fact that the organization gets the analysis back – which is the key thing for many organizations. Because many within the charity sector just don't have the data, analytical skills and capacity and capability to work with these quite complex data sets. So they get the analysis back and importantly, they don't get the individual names of people who reoffended or not. Because that data, that administrative data, actually stays with the government. It stays within the Ministry of Justice and doesn't come out. So it's a model that actually, we're securely sort of data sharing. But actually data sharing in a safe space. And we're not sharing individuals' details.

>> [Inaudible], that's great. So Rod, let's shift over to you, the organization where you work. The Prisoners' Education Trust is using this government data to try to reduce reoffending by giving prisoners a path to education. And you were there when this whole thing was started. So what's your experience been like, working with the Justice Data Lab? And how has it affected how your organization now collects and manages data? How much of a burden is this on your existing work?

>> Right, I think the first thing to say is that it's a problem that lots of organizations face that are working in the prison system. Or with people at risk

of offending. Is getting any kind of data on the real impacts that their work has. Very often you hear people talking about, we haven't heard of anyone that's gone on to reoffend since this program. But that really isn't terribly useful because of course the people that you stay in touch with tend to be your success stories. And you don't ever get some real proper analysis. So being able to track offenders who have been in prison after their release is incredibly important. That's a very difficult thing to do because their lives are often quite chaotic. There's no particular reason for being sure that you're going to be able to stay in touch with people. And the danger and the risk is that the ones that you are able to stay in touch with are your success stories. So it's incredibly valuable to be able to get some real hard-edged analysis which tells you what's happened to those people. The data that you actually need in order to submit to analyze your work in this way actually is really incredibly simple. I mean, all you really need to do is to have enough identifiers so that you can identify the person that you've worked with so that they can link with other data sets. So name, date of birth. In the UK we have a prison number which is also a very useful identifier. And all you need to know is the identities of the people that you've worked with and when you've worked with them. And certainly from our point of view, well that was information we were collecting anyway because we had to contact these prisoners, of course they were giving us their names and indeed their prison numbers so that we could be in touch with them. And we also, as it happens, asked for their date of birth. So we were collecting the core information that we needed already. Which is just as well, because of course one of the features about looking at reoffending is that you're always looking at cohorts that have left prison some time ago. You've got to have worked with the prisoners, they've got to have gone on to be released, and then they've got to be out in the community long enough to either reoffend or not reoffend. So we were in the fortunate position because of the way we work, we had that administrative data already. And were therefore in a very strong position to be able to submit it to the Justice Data Lab. And the results really have been very important to us. They've given us some surprises. They've confirmed the power of education and provided some really important evidence around that which I think is important to the UK context. And we're looking forward to being able to take this even further. I mean, Tracey described the link-up with the data on the national police computer database which talks about reoffending. But the MOJ, the Ministry of Justice have also been able to make a link-up with data about employment. And so we're looking forward shortly to being able to find out whether education also helps these prisoners get employment more than prisoners that don't have education. And that could be really important as well.

>> Right. So there's a lot packed into there and we'll come back to a couple of those pieces. We're already starting to get some questions about the actual data burden on nonprofits to interact in this environment. And you make a pretty compelling case that it's really not that much additional work; you already had the data. But let me introduce our third speaker, Justine. The Rhode

Island Innovative Policy Lab is using the same model or the data lab's model to help rebuild whole communities and to try to help ensure that prisoners are successfully integrated as productive members of society. Can you tell us a little bit about RI IPL and how does your model balance efficacy with privacy? And how do you see models like RI IPL perhaps becoming more common as an evaluation tool in civil society?

>> Thank you, Lucy. I'm glad to be here and share about what we're doing at RI IPL. RI IPL, I wouldn't characterize as a data lab; it's an innovative policy lab. And of course innovative policy depends on, and any insights depend on great data. Great data fuel great insights, fuel great policy and decision. And eventually fuel impact. So while we have data as part of our core, that really isn't our mission. Our mission is to use data and science to impact policy and improve lives. And RI IPL is a lab that I started. I'm a Professor of Economics and International and Public Affairs at Brown University. And a research associate at the National Bureau of Economic Research. And it's an initiative that I started a little over 2-1/2 years ago with foundation funding. And in partnership with the government of Rhode Island to use data and science to figure out, how can we move the needle on a host of important issues that have been challenging society for a very long time. And as a social scientist, one of the things that I've noticed throughout my career and that I've really worked to overcome is the fact that a lot of really great research is done and oftentimes it seems as if it sits on a shelf. And not only does it sit on a shelf, you know, it could have a lot of insights and impact for policy. But there really isn't that ability to take that research and translate it into action for policy. And maybe evaluating new policy changes. And as I've worked over the years with various governments using administrative data for particular topics and issues – for example, student loans, crises, privatizing social security, school choice, environmental regulation – what I learned is that there really is a need for having kind of what I would call a fully vertically integrated model. Since I am an industrial organization economist, I think a lot about the scope of a firm and the structure of a firm. And so what we set out to do is really make a kind of an all-in-one, all-encompassing from ideation all the way through to measurement innovation and measurement again and policy change, to put that all in one place. So at RI IPL, we start with our government partners by defining policy goals. We then have built a data resource so that very quickly we and our government partners can assess how current programs are dealing. Understand the key challenges affecting target population. And then use economics, behavioral economics and social science to generate insights that are fact-based and research-driven that suggest ways to improve. And then we can either take action if we think that that is the best thing to do, and change policy to [Inaudible]. Or we can try something new and measure what the impact is. And so we do this process over and over. It's kind of the foundation of how we work. And we apply this process in kind of four general areas. The first area are projects that fall under our goal of closing the achievement gap from cradle to college. So this includes things for adjusting gaps for early childhood health, early childhood education all the way through

college in the labor force and incarceration as our other two panelists have been talking about. Social program impact is our second broad goal. And that's improving the effectiveness and efficiency of our social safety net. So in the United States, many states face budget crises and have kind of growing tensions and obligations. We need to figure out how to spend smarter, not necessarily spend more. And to do that we need to understand where our dollar is being most impactful and how can we get more out of them? For all of our social safety net programs. Our third area that we're working in is criminal justice. And this is improving our criminal justice programs so that we can prevent crime and reduce recidivism. And our fourth area that we work in with our government partners is designing regulation that works. And what I mean by regulation that works is, most regulations have noble goals. For example, you know, preventing pollution. Or reducing greenhouse gases. But oftentimes those noble goals come with unintended consequences, or distortions that are caused by our regulation that we might not have thought about in advance. And so when we design regulation that works, we have regulations that both meet their goals and minimize unintended consequences or distortions as well. So we work through this process; we have dozens of projects ongoing with our government partners. And at the core is using a government administrative data securely in an anonymized fashion and effectively so that we can quickly address society's most pressing problems.

>> So can I just ask, since many of the people on the call are coming from the not-for-profit sector in the US, how does RIIPPL interact with nonprofits? Or how do nonprofits interact with RIIPPL?

>> So at the moment when we interact with a nonprofit we often are interacting with them in coordination with our government partners. So nonprofits interact with our government partners on a regular basis, whether it's in child protection, whether it's in social services such as the SNAP program. And so we often interact with them as we are working with our government partners to evaluate current programs, understand how programs are used by individuals, and also with designing and testing what we would call smart policy changes. Changes that promise to help make policies more efficient and effective.

>> Right; thank you. So that's a nice segue, then, Tracey, to – going back to the work that NPC does. Which is really, part of it is about inspiring nonprofits to take more initiative in measuring their own impact. And promoting the importance of measurement is obviously crucial. So improving access to government data is a part of this. But what else is needed? And how can nonprofits get better at using this kind of data? How can they hold both the government and themselves accountable for how the data are collected and stored and used? How does that set of relationships work now? And where do you see it going?

>> To address the first part of the question, when we're just looking at what role do charities have? Does the Justice Data Lab encompass everything? And essentially no, that's only one part of the evidence puzzle, really. And it's just so important to actually really gain the feedback from the people that you work

with to undertake qualitative research. So not just to collect information about, you know, did it actually work. But how did it work? And why? And for whom did it work better? And the best ways you can do that is actually when you use qualitative assessments and you talk with people. So that's one part of it. And the second part, when we look at sort of comparing about holding the government to account and charities themselves, I think we all have a role to play in being responsible with people's data. This is really important. And it has a lot of value to the person themselves. And, you know, the private sector will say it has a lot of economic value as well. So as charities, we do need to safeguard that. And also think very carefully about how do we use that data and assure especially for charities that we're using that data to actually support the individual, to help them in their pathways. And also that we're learning from that data to enable social good. And I think that strong compelling case is essentially what really won over the Ministry of Justice when we were able to present in that case that by actually, let's work together and see if we can develop the model. We understand that privacy should be at the heart of it, but we want to be able to ensure that we are always working towards social good. That we're not providing any services that might have some untoward circumstances. Circumstances that actually many organizations wouldn't know that they're achieving perhaps a negative impact. Because they just need that help in measuring. And actually as – sorry, just to repeat myself, there are just so many charities that didn't go into charities to be data managers or analysts. They want to do good. And so support to help charities in order for them to kind of see and, you know, measure whether they are doing good is essential for that.

>> That's interesting. So that's a nice segue; we're starting to get some questions in from the listeners. But I want to get two more kind of very different pieces of the puzzle out before taking one of those. So I'm going to go back to Justine and then to Rod. Justine, you're clearly an expert at managing lots of different data sets and bringing them together. I think I read that RIIPL's database includes more than a hundred previously independent datasets. So to Tracey's point that nonprofit leaders don't generally go in to become data experts, they don't go into this line of work. But you have that expertise, and RIIPL is built around that vertical integration as you described it. How does RIIPL in its current work assess and manage the quality control, if you will, across all those different data sets? And is there something that a set of organizations – a RIIPL or sister organizations if there are any – that they can do that eases that burden for either government or for the nonprofits or both? And you're also muted.

>> That's a great question. And something that we thought about a lot here at RIIPL. So one of the challenges – the government data are great. And administrative data are really great in a lot of ways. Because unlike survey responses, you get measures of what people actually do or have done as opposed to what they say they do or want to do or have done. And so in many ways that's wonderful. A great example of this would be voting data. If you ask people

whether they voted in the last election, it turns out many, many, many more people voted in the last election that actually did vote in the last election. And so oftentimes surveys can be difficult. They can also be difficult because people have problems recalling things that they've done in the past. And so, you know, if we wanted to look at what people did six months ago or what happened in a child's life six months before they took an exam or a standardized test that may have impacted that test score, it might be very difficult to get that information from parents – and more importantly, from a child. So administrative data are very powerful in that sense. However, administrative data are typically made for keeping the trains running. So they may be in case files, or case management files. So many of you have gone to a doctor's office and you know that the receptionist or the nurse is interacting with a Dewey interface where they're pointing and clicking and typing things into boxes. And all of those data then get put into some back-end database and the way that that database is structured is for accepting input, storing them, recalling them into a Dewey interface. Not necessarily for answering the question, you know, what was your total health bill? You know, someone's total health bill over time, how does it trend as someone ages, right? For that, you actually might want some really easy summary statistics on health bills and age. And it turns out it's a very complicated set of procedures to go through to get case management files into something that looks like age and total health spending. And it's even more complicated if you wanted to know age and total health spending and incarceration. Because now we may need to know – have many different case files and case management files and we need to get them into a shape where we could actually get some insights out of them. So at RI IPL, what we've done is built in anonymized and secured data in a relational database across these different agency records. But we're able to join records, these anonymous records with an anonymized ID. And then not only have we done that for all of these case management systems, but we've built critical-derived tables on top of that which take these case files and turn them into very usable formats for insights. So that we can do something very easily, for example, plot incarceration outcomes by characteristics of students while they were in public school. So that seems like a very complicated thing to do. But in fact, we have – at RI IPL we run the data-drive policy track within our Master's in Public Affairs program. These are students who would be going out into the NGO world to work in many not-for-profits in policy and also in government. And they are able to, with just a few computer science classes, sit down and make graphs of trends in, you know, in the relationship between incarceration and school outcomes within a week. Which is amazing, actually. And the same for our government partners, as we are working with them to be able to access and utilize the theater resource. It is designed in a way that it takes all of the complicated, difficult administrative data sets and makes them so that they are in a format that somebody can actually extract insights from very quickly. So that's kind of part of the center of what we've done, or what we strive to do. And the reason is, because remember our goal was in a timely fashion, go from policy problem to great research to policy solution. And that has to happen

quickly. Not on like an academic timeframe. And to do that quickly means that when that problem comes in, we need to ideate, we need to think, we need to come up with a plan. And we need to execute that plan. And we can't be spending a year or two to wrangle the data or understand it. So in this way we have done that wrangling all at once. And delivered something where people can really start generating insights quickly. And that's been great to see. And as I said, we're just a little over 2-1/2 years old and this half-year we've really been able to capitalize on our investment. And are really moving at an astonishing speed along with our partner, which has been really wonderful.

>> Great. And you're touching on really helpful detail on the nuts-and-bolts of the data management. But also beginning to hint at a question I want to ask Rod about specifically. And then our audience has two follow-up questions about this point. Which is really about kind of the culture change and skill change that having access to this kind of resource demands or makes an opportunity for both people entering the nonprofit sector you – Justine just talked about the way it changes the nature of thinking about these issues in the academy, the ability to integrate it into coursework. I mean, we're seeing this –

>> And if I can just make a quick plug for the Watson Master's in Public Affairs data-driven policy track program, if you come, you will both learn policy and a bit of computer science and econometrics. So you will be ready to go out there and work with Rod and Tracey and Lucy and all of our listeners to help integrate data into their decision-making. I will turn it back to Rod.

>> Now, so let me just transition the question. And that's absolutely fine. Because we need to – the need for people who can speak these different languages, if you will, is obviously growing. But, Rob, on the culture question, if you will – or the sort of assumptions and expectations of work in this age question. I want to ask you about what it's like and what it means to be then taking the initiative to have your results assessed against this larger database, to be transparent about that so that folks can actually – you can hold yourself accountable, the government can hold you accountable, the people you work with can say hey wait a second, what's the effect? What is it like to operate with that level of visibility? With the level of visibility that data makes available? The transparency issues, how does it change how work happens at the staff level? How does it change the way your Board and trustees engage? It's a pretty significant new set of expectations.

>> Yes, yes. First I'd just like to comment a bit on what Justine was saying. And just make the point that although the Justice Data Lab is focusing on the not-for-profit sector and the work that it does, actually there is a huge amount of data out there which government is in a position to be able to mine for some really important information without involving the external agencies. And I think there's a big agenda for government to really make the best use of the data that it's already got available. The point about the impact on an organization in thinking about whether to sort of deploy a tool like the Justice Data Lab, I think is a very good one. I think there are certainly potential risks

from the point of view of a not-for-profit in going into this. There are many reasons why you may not get the results that you confidently expect from this kind of analysis. There is an issue to do with – there’s one big issue which is, are you going to get new results which are statistically significant? And that’s a product of two factors. One is effect size. So if the effects of your intervention – even though it’s terribly valuable, is only two or three percentage points, then that’s going to be more difficult to show out in a statistical analysis. And the other is sample size. So if you’re only working – I mean, the Justice Data Lab works with a minimum number of 60 participants. But for a number that small, you really do have to have quite a large effect size for it to show up as statistically significant when compared to a control group. So that’s a problem. And it is an issue for a charity to think, well, we tested this but we’ve now got published data out there which says that you can’t – that it doesn’t prove to have any positive effect. And so being able to – being sufficiently sophisticated to understand that just because it’s not statistically significant it doesn’t mean that it doesn’t work. I think it’s also really important to understand – I mean, you do need a level of sophistication in understanding how the technique works. So that you can interpret the results sort of sensitively and intelligently. I mean, for something like ours, which is an educational program, an educational program which is delivered in really high volumes, because for a not-for-profit we work with thousands of people over the course of the year rather than tens. Actually, the Justice Data Lab is a very good kind of avenue to go down. Because in a sense, people that are engaging in a program are already buying into some positive future for themselves. So they’re likely to be a more positive group than the group that don’t engage with an education program. And so in a sense it’s not surprising that you get some positive results for that group. And that is something that we’ve been able to test by looking at people that for various reasons we turned down from being given course. And sure enough, just the fact that you’re interested in applying and interested in engaging in itself is correlated with a positive impact on your future. So in a way we’re a prime example of an organization that’s going to do well out of that analysis. To contrast that, if you’re an organization which is trying to work with some of the most difficult, complex needs, multiple problems and in a sense if you’re trying to kind of cherry pick the most difficult clients to work with, when the Justice Data Lab creates a matched control group, are they going to be able to match people that have that same level of difficult situation and complexity of case needs to address? So there’s a risk that actually you don’t get a positive result. Not because your program isn’t great, but because by its very nature you’re focusing in on the hardest to help. And it’s very difficult to have a matched group that is sufficiently well attuned to pick up on people with that level of problem. Even though this is something that the Justice Data Lab is working with and trying to improve. So as with all data, all data is absolutely fascinating. They give you some wonderful kind of insights. But what it tends to do is raise an awful lot of questions. And you do need to be sufficiently sophisticated, I think, to be able to understand some of the issues involved with that and to be able to interpret what’s coming out of this without jumping to

some conclusion. And you need some space in order to be able to explain that to a wider public. Because it's very easy to see this as a kind of black-or-white, is it positive or is it negative kind of answer. And you do need that level of sophistication to understand it. And you're right, we did need to go to our Board to say we think this is an excellent analysis to conduct. But you really want to know that the results could turn out to be negative. In which case actually we've got a problem. To their credit, the members of the Board said, well, actually if we've got a problem we need to know we've got a problem, because we ought to kind of be thinking about whether this is a proper use for our charitable resources or not. So it's a question we want to know the answer to. As it turns out, of course, education is a really important factor in helping people to get their lives on track and have positive outcomes after release. But there is always that possibility that it doesn't show up that way.

>> And just to add to that, we've learned that there have been some charities that actually haven't used the Justice Data Lab because of that fear, that a risk of failure has actually been a complete barrier. And they've actually said to me that we've actually got the data but we can't submit. So that's why sort of the poorer organizations like Rudd's and others that have used it. Because it's almost like you have to be quite fearless in the fact of saying, well actually we're prioritizing that we need to know whether, you know, how effective this program can be. And just to draw on another point as well about the Justice Data Lab. To their credit, I know that I'd really like to say it's really great work that the Ministry of Justice did. But they're continually always improving the model. And so at the beginning they couldn't even – they actually said to organizations that worked with complicated people like people who use drugs or have had mental health issues couldn't use it. But actually they bring in more data sets so that they are able to identify within the comparison group. So now they can do analysis for those people that have used drugs, or have mental health issues. But that doesn't negate – there's always going to be issues when you do these match control group analysis. Just the same as you have issues when you do randomized control trials, say. It's exactly what Rod has said; it's partly about an education, why the piece needs to take place with the charity sector. And we've tried to do a bit of that at NPC, but there's certainly much more to be done. And especially that expectation that not everyone is going to have a 30% reduction in reoffending. Most of the studies we saw was actually sort of like tinkering around like a 2%, 2 to 5% reoffending. And actually many charities have often claimed before, you know, 50% reductions. And so actually we need to kind of change that mindset where actually it could be much smaller than expected.

>> I mean, given the costs of reoffending, actually a few percentage point reductions is really worth playing for. I mean, it's going to make a huge difference to society. And so the costs that come with crime. And when you work in social policy, you really ought to be thinking in terms of single figure percentage point impacts. Because that's the kind of – realistically the sort of scale that different interventions are likely to have. Particularly if they're relatively cheap

and relatively small scale.

>> Yes. So, great conversation here. I want to bring in some of the questions from the folks on the call that I think align with kind of the broader context that the three of you have been drawing. So I'll read out the question and then look to each of you for a brief response if you have one. The first one speaks really to a kind of a very large-scale potential cultural shift as your work becomes more common, if you will. Or the use of data in this complicated methodological ways becomes more common. So the question speaks to whether or not there's actually – we're experiencing or we're about to experience a re-shaping of how human services work is delivered in ways that preference digital platform delivery precisely because they kind of automatically generate some of the data that's needed to facilitate evaluation? The person who asked the question speaks that this seems to be happening in the ed-tech sector, in particular. So the issue here is that the ability to do this kind of data-driven methodological evaluation and comparison actually then possibly intentionally or unintentionally pressures the system as a whole to focus on delivering services that can easily be counted. Or that are delivered digitally and so are automatically delivering the data – ed-tech, telehealth, teletherapy – in a way that's kind of privileging what might be service delivery decisions over data collection possibilities. And I'm just curious if anybody on the call has a thought on that? An observation? A concern?

>> I mean, I would say that – I mean, as I said earlier, in order to run something through the Justice Data Lab, really all you'd need to know is who are the people that you've worked with. And the dates when you did it. And being able to identify that. I would be surprised if the main driver for people capturing data digitally is ever going to be a desire to run data matching exercises with government data. I think generally speaking people capture data for their own administrative purposes, and it's that that gives it a lot of its power. I think the other thing to say is if you are a very small organization dealing with small numbers, then actually you're probably not going to get fantastic results out of a big data matching exercise. Because your sample size isn't going to be enough to demonstrate the value of what you're doing. But what you can do is piggyback on other research which shows that programs of the kind that you are running do have that scale of impact. And then you can focus on just needing to prove that you're a good example of that and that the kind of intermediate outcomes in terms of achievements by the people that you're working with and their own reaction to the program that you've delivered is positive. And that should make the case that, given the other evidence that's available, that you're doing the right thing and it's going to be producing a positive end and effect.

>> Yes. And I'd also say that actually it is about proportionality; we don't want lots of small charities feeling like, oh my goodness, tomorrow because we're in a data-driven world we all need to collect data. Really if you have a sort of clear theory of change about how are you helping organizations, it could be that your actual ultimate aim is more like an intermediate outcome that supports somebody to stop reoffending. Or to improve their educational qualifications.

They are some of these bigger aims, in fact. And so actually the soup kitchen-plus that provides a safe, warm space for homelessness, for example, shouldn't be thinking about yes, we stop everybody that comes through our soup kitchen. Just not homelessness, but actually they might be thinking, our role here in the wider kind of ecosystem of supporting people is providing that safe, warm space of empowering people so that they next move on to the organization that actually provides that intervention that helps them to actually get on that way. So I think there's a bit where on the charity sector, we really do need to advocate and talk up for ourselves about, you know, where we're best placed to do some of these things. And some of it is right at the hard end of like getting people to stop reoffending. And others are sort of, you know, a little bit middle or low, the first rung stats of helping people on the journey. All are equally valid.

>> I would probably say from my point of view that, you know, a little bit in agreement with Rod that, you know, a lot of the administrative data really is measuring outcomes as opposed to delivery method. So for example, we might be interested in whether somebody is re-incarcerated. You know, there's nothing in delivery service that's going to change that as the case file; you've already generated a case file. And using that helps us understand what outcomes are after programs are administered. On the other hand, I would also point out that to some extent the world is pushing us towards more electronic delivery as opposed – not research or evaluation, but actually just technology. So two good examples of this that might be helpful for the NGO world are first, it is the case that most government agencies that do traditional surveys are now having a very difficult time getting phone surveys accomplished. Why? Because people don't have land lines. They don't answer their cell phones. They do answer text messages, actually. And they would prefer to do text messages online than they would – this is for very low income people – than they would asked it on a phone for a period of time to answer questions. So taking advantage of that might mean that an NGO adopts an experience sampling method, which is a survey that kind of sends you one question, you know, every – like twice a day. And if you answer that question in the moment, you accrue 25 cents on a gift card that can be redeemed electronically. It turns out people are very willing to answer those types of questions. Much more so than a long phone survey. And you can collect the equivalent of a phone survey over the course of a week with these little bits of survey questions. They also are really great for things – Tracey mentioned soup kitchens. We've tried this out with measuring hunger. Traditional managers of hunger and food insecurity are actually very backward-looking. They ask you how many times you've experienced something over the past six months, maybe a year. You can actually ask this in the moment, "Are you hungry? Are you worried about where your next meal will come?" And the nice thing about administrative data is that if we also ask somebody, did you use your EBT card today? And, how many times did you use your EBT card in the last month? We actually can measure how good their recall is. Right? Against administrative data. We learn that recall actually is not very good at all. And that maybe measuring hunger in the moment is much better than asking people

retrospectively how often they were hungry. And then thanks to technology, people would rather answer the questions in the moment anyways rather than do a long phone survey, it turns out, these days. So to some extent it's hard – it's getting back to Rod's discuss of correlation and causation. It may be hard to know if the need for evaluation is pushing us toward more electronic delivery of services. Or if it's pushed by demand. This is actually how people want to interact with their services. And it's a nice byproduct that we get better information and can make our programs more efficient and effective at the same time.

>> Interesting. It's a great set of considerations. The way the question is shaped, I'll just read a little bit more into the question to add another possibility, which is just the economics of digitized services at one level seem a lot less expensive to funders. And so that's just another factor in this dynamic. I think the points each of you has made, though, is a really important set of considerations about both external forces of change and the culture of work within the space and to be aware that there are tradeoffs. And the next question speaks specifically to tradeoffs, so I'm going to meld a couple of questions together here in the interest of time. We only have seven minutes left, so I'll ask you to try to get to this. But the question is, what are the tradeoffs between approaches that focus on adapting existing data and academic data-sharing infrastructure to make it accessible to nonprofits and creating new data sharing infrastructure that might be more tailored to the needs of the nonprofit sector? And your two examples give us a little bit of both. The other question I'll tie to that is, as you're thinking about that, to where does the ability of your local or municipal or state or national government and its own ability to make this information available, how does that fit in for nonprofits or NGOs that might be wanting to engage in this way? So, tradeoffs in the models. And then any advice for nonprofits working with governments where the capacity issue to actually manage the data is it sort of seems nowhere resident in the ecosystem.

>> And I think I would say that the main reason why governments ought to be doing data matching is to understand what government does. So the really important reason why you should be matching prisoner data with reoffending data is to understand what you're doing with prisoners while they're in custody to see what impact that then has after they're released. And the Ministry of Justice in the UK has done some very interesting studies in that area that have been frankly deeply embarrassing, but caused them to rethink some of the programs that they've put in place. Once you've done the heavy lifting of creating the mechanism to match the data and there's huge numbers of problems about data cleansing and getting the matches to secure so that you know they work – once you've done all that work, actually it's not that difficult to then ask some question about relatively small groups of people that have worked with some of the not-for-profit sectors. But actually government understanding the potential of its own data, I think, is the right place to start. And I think that is what drives a lot of the baseline costs in doing this. And quite rightly, because they ought to be marking their own homework before they start trying to mark

ours.

>> And just talking about the different models, I think the benefits that you can have with the Data Lab model is where you have a number of organizations that share – that want to answer the same kind of question, do we stop reoffending? Do we help educational qualifications? I can have many different types of charities and they all have that common type of outcome that you’re all trying to achieve. So actually by having a team within the government department we’re able to kind of routinely provide this. You know, the same question over and over again. That’s really important. But at the same time, it’s also important that we have – like in our policy labs like Justine’s, so that we can put some big policy questions, you know, over to that side. You can do some more sector understanding and drill into some questions. So I don’t really – I guess within the economic world there was always a tradeoff about where is money spent. But I would like to see both of them, in a way. But I do think it’s important that we do provide not-for-profit organizations with access to the data and that is analyzed. And actually it’s not just nonprofits. So what the biggest reasons – one of the interesting things we found about the data labs was that government departments, actually public services were using it. Prisons have used it. All organizations that I thought, don’t you have access to this data? You’re the government. Well actually, maybe they do have access to the data, but they don’t all have statisticians in prisons that are able to turn this data around. And so it’s been used by the private sector as well. So I think where you have some kind of common questions, I think for economic and efficiency, it might be useful to just have a team that’s able to pull this out. But also at the same time when we have bigger researcher questions, you might be looking at need, not just about impact, then this is where it’s our policy labs where it’s taken to place. And that’s where I think it’s also important that nonprofits are able to raise some of the questions to these policy labs as well.

>> Justine, did you want to jump –

>> Yes, sure. I think, you know, I totally agree that both having government use its own data and also allowing vital not-for-profits that exist in society and in this ecosystem as addressing deep issues of poverty, that they’re also able to access data. This is very important. And that’s part of what our goal has been at RIPL. In the sense that we wanted to make a data resource that is usable for quick insights. And we do actually have – have started taking in data from some nonprofits in partnership with our government partners where those are important additional data because these not-for-profits are providing a service that we really want to understand as part of one of our critical projects. The goal and what we’re working towards is that government is actually able to access the data in the same way we are. And that the data are in a format that somebody with a data-driven MPA can analyze them; they don’t have to be a statistician or an economist or a computer scientist. Because they’re in a usable form that is optimized for insights. And so creating this data resource for the state government allows them to use the data and eventually can then

allow them to provide those things to the not-for-profits very easily. Because it doesn't take, you know, a PhD in Statistics or two years to wrangle the data in order to deliver something. And in fact, you know, we were having this conversation just the other day with one of our largest agencies. There's a nice kind of ecosystem of analysts across agencies that we're putting together to learn how to use the database and create that institutional knowledge. One of the things that came up was a comment about, you know, managerial impact of this new resource. If somebody asks me a question, instead of saying, well, I can give you a not great answer and just giving you a not great answer, I can say, "I could give you a not-great answer now, but if you waited one week, I can actually deliver a really great answer." And maybe one that wasn't possible at all before. And as they learn to use the database and as we continue to develop over this next year this integrated approach, I think that it will be easier for program managers and directors of agencies as they cooperate with the not-for-profit to put those data together so that everybody is learning and able to have the greatest impact per dollar spent. Our other goal also at RIIPPL is that we make this resource available to other states. So we've already started talking with other states. The process that we've generated is actually scalable to additional states. And so my hope is that, at least here in the US, this is a model that scales and increases its impact to facilitate fact-based policy both in the not-for-profit and the government sector. So that we can actually really put a dent in these problems that have been facing us for so long.

>> Great. That's a fabulous place where I'm afraid we're going to have to end this. I just want to say that just looking at the stream of questions and comments, there's lots of concern and additional insights and questions about how, when we put kind of data in the middle, at the very least what's happening are some new kinds of relationships with new kinds of skills and new kinds of incentives between governments, nonprofits, funders, and the commercial sector. And that's, you know, that's the world that each of you is working in and your work is to some degree representative of. So I apologize to the folks on the call that we're not going to be able to get to all those questions, but we've got as many of them as we can here. And maybe this is an example of a topic that needs yet another conversation, which the lab would be delighted to host. That wraps up our time today. Let me thank our panel, and thank you for participating. If you did miss any of it or you want to share the conversation with your colleagues, it will be available on digitalimpact.org, and the Digital Impact podcast on iTunes. If you want to learn more about NPC's data lab work in the UK, see thinknpc.org, and you can learn more about NPC's work with the governance labs at medium.com/data-labs. You can learn more about Rod and Prisoners' Education Trust at prisonerseducation.org.uk. You can learn more about Justine's team and the Rhode Island Innovative Policy Lab at riipl.org. And of course please be sure to check out digitalimpact.org and the Digital Impact Toolkit for more opportunities and a host of resources for advancing the safe, ethical and effective use of data. If you enjoyed today's conversation, be sure to join us for more in the coming months. And I'd like to encourage you

one more time, if you want to continue this conversation, if you want to frame it in a different way, if you want to bring up another topic, please be in touch with us at hello@digitalimpact.org, which is where we get the ideas for these roundtables. I'm Lucy Bernholz at the Digital Civil Society Lab at Stanford PACS. And thanks, everyone, for joining us. Bye-bye, now.