

TRANSCRIPT

"HOW CAN WE USE DATA FOR SOCIAL GOOD?"

A Conversation With Jake Porway

Moderator: Elizabeth Eagen

ANNOUNCER:

You're listening to a recording of the Open Society Foundations, working to build vibrant and tolerant democracies worldwide. Visit us at OpenSocietyFoundations.org.

ELIZABETH EAGEN:

I'm Elizabeth Eagen from the Human Rights Initiative and the Information Program. And I'm really pleased to welcome Jake Porway from DataKind. DataKind is an organization focused on providing opportunities for data scientists to work with the non-profit world and to work with data for the greater good, as you'll see from Jake's title and his talk. (COUGH)

Jake-- most recently was a data scientist with *The New York Times* and he's been-- a great partner for us in thinking through what exactly (COUGH) we can do with large amounts of data. What can N.G.O.s do when they have-- when they have a master data but they're not sure what it can do for them.

And what can-- and how can they think through asking questions of the information generated by the day-to-day work of an N.G.O. To make it do more for you when you're-- when you're running a project in a civil society. And go ahead, take it from there.

JAKE PORWAY:

All right. Well, thank you so much for that intro. You know, Elizabeth has been an amazing partner for us because in doing this work, you know, we come more from

the data side, as she mentioned. That's been very powerful. But, of course, to do this work really well incredibly good partners understand how this could be applied. Or know the situation on the ground.

So I'm really excited to be working with Elizabeth and O.S.F. and getting to come talk to you guys. So thank you for coming today. So-- as Elizabeth said, you know, I'm-- I'm actually more excited about the conversation that we're all going to have.

But so to set things off, though, I thought I would just sort of lay the ground work of some of what we do. And before I get into what DataKind does, I-- I feel like it's always helpful to think about what data is. Really set the stage for that.

And so I'll often ask people to actually think about data. Like, picture data. Like, you know, let it-- actually take a moment. Just picture it. Like, let it wash over you. I see people closin' their eyes. Like, yeah, so-- so what do you picture when you hear data? Did anyone-- just shout it out.

VOICES:

Spreadsheets. (LAUGHTER)

JAKE PORWAY:

Spreadsheets. Exactly. Probably spreadsheets. Yeah. Anybody else picture anything else?

MALE VOICE:

A pile of documents.

JAKE PORWAY:

A pile of documents. Yeah, like, you have to document servers. Or even worse than that what a lotta people often picture is just the anonymous tube of numbers. Like, the matrix, right? And-- and that's a bummer because I-- I-- I even think of this this way when I hear big data.

And-- and that's a shame because data's so much more personal than that. And it doesn't take much to realize that except to think back to a time before big data. So if you go way back to, like, the dawn of humanity, like, way back in the Dark Ages, back to, say, like, 2003. (LAUGHTER)

When (UNINTEL PHRASE) like to get a movie. Right? Like, remember this? You didn't know if it was good or bad. No recommendations. You just kinda wander around. It was horrible. (LAUGHTER) Dark time for humanity. But thankfully,

right, we're livin' in 2014.

Where with the click of a button you can see customized preferences for your movies. Netflix knows exactly what you wanna see. And you can make better decisions about how to spend your Friday night. And that's awesome because we live in a world where there's now tons of data.

So Netflix collecting massive amounts of data about what you like. What other people like. It's crunching down those algorithms to help figure out what you wanna watch. And of course it's not just movies, right? It's how we-- how Google maps figures out how we get from Point A to Point B most efficiently. It's how we figure out what we wanna buy. (LAUGHTER)

Or even how we have-- relationships with friends and how our social networks grow. So this is this new age that's occurring because as-- as we move our-- our lives online, every interaction we have with each other, with our world, it creates data. Kicks off a little piece of a record.

All of that can be used to understand more about our world. The-- I mean, of course, you look at these examples, right? It's buying stuff. You know-- watching movies. Like, those are good pursuits. I certainly use Netflix all the time. But there's probably more we could be doin' with this data. There's probably-- a greater good we could look to.

And so to set an example of what we could do with this data, I always turn to a luminary of our times. Somebody who sort of really understands this world-- the great Kim Kardashian. (LAUGHTER) Of course. Of course. Now, I know a lotta people are like, what-- I don't really un-- you know, why does Kim play a role in this? And-- and the way that she fits into this story actually-- starts with the C.D.C. (LAUGHTER)

The C.D.C.-- everyone always laughs at that. I don't think it's the reason you think. (LAUGHTER) It's not as disease-- it's not a Kim Kardashian disease connection that they're probably imagining. But the-- the-- this story comes from the fact that the C.D.C. (PHONE RINGS) is always on the lookout for the next killer flu. The-- you know, the pandemic that could wipe out humanity. Oh, is someone calling in?

ELIZABETH EAGEN:

It's the C.D.C.

JAKE PORWAY:

Yeah. (LAUGHTER)

(OVERTALK)

JAKE PORWAY:

Don't approve and that's fine. Yeah. So data. They-- somehow they got the data stream. They knew. So these guys, they're always looking out for the next killer flu. The way they do that is they put people in hospitals and clinics and they record everybody's symptoms. And the second it looks like there's too many flu symptoms out there, they put out an alert so that they can quarantine or-- or (UNINTEL PHRASE).

And that's great. Except it takes a lot of people a lot time. So they wondered, is there a better way to know when people have the flu faster? And this is exactly where Kim comes in because no matter what you think of what she says on Twitter, she's really good at telling you (PHONE RINGS) when she has the flu.

There she is in bed with the flu (COUGH) all day, praying it goes away fast. And of course lots of people across Twitter are constantly complaining when they have the flu. Or saying that they're home sick. So a couple of grad students said, "If we just mined that information, maybe we could get a better read on this." And they were able to make a very boring but very important graph which shows that if you look at people mentioning flu symptoms online, you lead the C.D.C.'s predictions by about two weeks.

In two weeks faster you get real time information. And if you even link this up with where those tweets are coming from, you can make a visualization of flu season over the U.S. in real time. Which is either awesome or disgusting, depending on whether (LAUGHTER) (UNINTEL PHRASE) the virus.

So this is an example to me of how all that data that's kicked off from something as seemingly-- superfluous as Twitter could be used for the greater good if you really just knew how to apply it. And what I was really excited about in my past life, as Elizabeth was mentioning, is that I was a data scientist. Which-- meant that-- basically it's nerd. (LAUGHTER)

But really means, like, a statistician or-- or-- a computer programmer who can work with data. And what I loved about being a data scientist was that we were an enthusiastic bunch. Data scientists don't just spend their time working on data 9 to 5. No, they're goin' and doin' side projects at night.

Downloading government data and just playing with it. Or increasingly going to hack-a-thons on the weekends. So for those who aren't familiar, hack-a-thons is 24 hour, 48 hour geek fests where data scientists, developers get together and try to figure out the-- what cool stuff they can build.

And I remember being really excited by my first hack-a-thon because when I went, oh, man, I'm sittin' next to the most amazing machine learning expert. An amazing developer. Like, this is how we're gonna make change with this data. We can build whatever we want. We can make a difference. We can do stuff that's gonna have so much impact. It's gonna be so different.

And the stuff that we came up with was so unfulfilling. I mean, we came up with

apps to help park cars. Apps to find local deals. I mean, it was great but it was more of the same. So there's this great resource of data scientists just spending their time making Twitter for pets. (LAUGHTER) That sucks. (LAUGHTER) It's so disappointing.

But then on the other side, what's really cool is to me, you know, people talked about what does big data mean? Well, I think one of the important definitions or-- or interpretations of big is that it's expansive. So it no longer belongs just to tech companies. But now groups like this clean water N.G.O. are inundated with data. And I don't just mean survey data that they're digitizing. Sure, there's that.

But there's also data that the government is opening up about well locations, the satellite imagery coming now-- to our computers that you can download. I even just saw-- a new device that called water canary. And if people haven't seen this, it's a small sensor you put in a well.

It tweets every minute about all-- tons of different indicators of the-- quality of the water. So how-- so this is amazing. Now all of a sudden your little clean water N.G.O.'s basically a tech company, real time data. And that's exciting.

'cause they can start using those same applications that we used to help buy goods and watch movies. Except that they have-- no one helped them do this. Right? I mean, I-- I don't know how many non-profits you know can afford or even knows how to ask for data scientists. But there's not many-- that I know of.

So we thought this was a great opportunity. Wanna get these scientists who are already working on great stuff-- who are spending their free time workin' on not great stuff to work with people who now this data to tackle really interesting problems.

And that's how we formed DataKind. Non-profit to build collaborative teams. Not just being data science for the social sector. That's so hubristic. The idea is more to build collaborative teams of people who have the problems and have the data.

Bring them some data science resources so that the data scientists can have social impact. Social organizations can maximize their impact. We do it in a couple of ways that we'll-- get into more detail on later. But the basic idea, again, is building these teams to work on these projects together.

So I'll show you a few examples before I close. One comes out of D.C. and these little non-profit called D.C. Action for Kids. A four-person non-profit that has set out to look out for child well-being in the district. And they were really excited because the government opened up all of these data sets about child well-being.

Educational scores, parents' income levels, food availability. And they said, "This is awesome. We can look at this problem in way we've never looked before." Then they remember data looked like this. Spreadsheets. And this is just one that's, like, hundreds of P.D.F.s and C.S.V.s.

For a small non-profit, they had no idea how to take their first step. So we teamed them up with some data scientists from *The Washington Post*, The World Economic Forum. And together they turned all these spreadsheets, this, into this. This is an

interactive map of child well-being.

And what's great about this is that you can interact with this to understand at each level-- how different indicators change over the district. It's nice because at a high level you can get executive summaries of your policy maker. You can read the firm's findings.

If you are a concerned citizen, you can mouse around and actually look at the information itself. Or if you're really nerdy, if you're really into you can download the data that's now all been combined at a neighborhood level, not census-tracked or ZIP Code, into one document.

So you can actually do that analysis yourself. And what-- what's more important to me about this was maybe not the findings from this alone. But the gasp that I heard when they unveiled this to the rest of the other partners in the Kids Count network. Because to me I look at this and go, "Oh, it's like a map."

You see lots of interactive maps every day. It's not that great. But they have all of these in P.D.F.s up until now. And just for people to see that this thing could be done actually has prompted the mayor of D.C. to come out and say, "This is how we have to start looking at child well-being.

"We have got to start using these resources." And now other Kids Count programs around the nation are saying, "What-- how do we do this? Let's adapt it for our work." So that's-- that's really exciting to me. A small little non-profit could use these tools.

Another example comes from New York City right here-- with the Parks Department. And they have tons of data about what they call the urban forest. Which I think is such a lovely name for New York's trees. And they have data about when those trees were planted.

Every maintenance event that happened. When they were pulled up. And they say, "You know, we've got this problem where one of our programs is to prune these trees. We go around, we cut down tree limbs that look dangerous, they don't fall on people and cars.

"But we kinda do it by gut. You know, we see a problem and so we cut it down. But we don't really know if we're making a difference. Or if we're doing it the most effective way possible, 'cause we don't have a lotta resources for this. So could we figure that out?"

So what's cool is that they teamed up with this guy, Brian Dalessandro. He spends his days at a marketing firm in New York. He uses data to understand if I show you an ad, do you think go on to buy the product afterwards or not? And what the nice thing about that is that is exactly the same statistical model that you can use to answer if I cut down a tree limb-- tree limb is there then an emergency later or not?

So he could take the same advanced analytics and statistics that he was using in his day job and apply it to this problem. And he was able to build not only this map that showed-- sort of an inventory of their trees, which is the first time that they had

really seen this. And I show it merely 'cause this is a talk. So here's something visual. But what's more important to me was they came up with a number: 22. Now, 22 is-- there were 22% fewer accidents, fewer tree emergencies on blocks where they were pruning than similar blocks where they hadn't pruned. And so this was mind-blowing to them. Not only were the programs working, somebody computed that number. No one had ever computed that number. And then other urban forestry programs from around the country were saying, "Oh, how can we apply this to our data?"

"How can we make better decisions based on this data that we have?" So that was-- that was really great. And I'll close out-- with one last example from Give Directly. May have heard about these guys. You know, just because you're in this world or-- they've been featured a lot in *The Times* lately because they give cash transfers directly to people by cell phone. They don't use micro loans.

And one of the ways that they want to (COUGH) expand is to find the poorest areas in Kenya to-- expand their programs to. And the way that they do this is they go into the villages. They go house by house and they look for the type of roof that each house has.

The idea is that if you have a thatched roof, has to be replaced a lot. That's usually people who are a little bit lower-- income level. Whereas people with metal roofs, something that's m-- a little more sustainable. It's-- it's one way of doing poverty measurement. Or-- I should say-- estimation, to be proper. But it takes a lotta time, just like the C.D.C. case. And so what they realized was that if you look at Google maps and Google images, you can actually pick out thatched or metal quite easily.

So here are a couple of thatched houses in the center here. And then over in the middle you can see a very clear metal image. So we teamed them up with some people from I.B.M. and Mozilla. And they wrote a program that-- ignoring what some of these-- some of these, like, Predator style looking (LAUGHTER) images that actually-- results in an algorithm that can pick out the locations of the roofs in these images.

So they can find the centers of these roofs automatically. And from that they were able to run it over the whole set of Google maps-- and create this basically-- poverty estimation map-- for this region in Kenya. And this was great 'cause people hadn't been able to do this with this granularity before. And it took about two hours to run the algorithm, all publicly available data.

It normally would've taken Give Directly a hundred days-- a hundred days of manpower to go out and do this. So think about that. Before they launched that program, they're budgeting in a hundred days, three months to go out and do this. It's done within a day.

Now, they're gonna go and check these. They're gonna double the-- do not trust, you know-- I wanna be very clear that in building these collaborative teams, we never just let the machine decide. (COUGH) But it could help target them into areas that they wanted to see-- to see, "Where should we start first?" So there are many more

examples that I could expound on. But I think, you know, these examples to me are-- are really illustrative-- of a concept that we've been talking about lately that I think's really important to understanding how this whole conversation around data is changing.

And I'm gonna borrow a line from my buddy Moritz Stefaner who's been talk about this a lot. And he sh-- pointed me to a book called *The Macroscope* which was written by a guy in the '80s. And he was saying, "You know, in-- in the old days we invented the telescope. And this let us see things that were huge, infinitely (UNINTEL PHRASE)." Infinitely far away. And that was awesome.

Then we also (COUGH)-- invented the microscope and that let us see things that were infinitesimally small, infinitely close, infinitely small. But what we really needed beyond this was something that was like a macroscope that would let us see the infinitely complex.

The patterns of society and nature that you can't just observe just by being in a room. And I think we're finally reaching that moment, thanks to data, which is now kicking off from all over these natural and societal interactions where we can start to create that macroscope.

Where data is no longer about reporting. Or about metrics. But it's about a new lens into the world so that we can understand and make better decisions. So I will-- maybe pause there and open it to discussion-- to say-- our sort of future plans are to keep this momentum going.

We've seen this project working fantastically and people want to do more. So our quest is to basically scale out to the point where we can provide these services for more people. We've had thousands of people sign up to volunteer. Hundreds to do projects. And so now we're-- really excited to get to that point where we're actually gonna start working on those so that we can start making-- better decisions. Not just about how we watch movies. (COUGH) But better decisions about the kind of world we wanna see. So thank you. And let's pause there.

ELIZABETH EAGEN:

Okay. So I'll open the floor to questions. I'm not gonna take moderator privilege. Please feel free. What's on your mind, spit it out. (UNINTEL PHRASE).

JAKE PORWAY:

And if it helps, I had a few points on-- like, I'll just say I had some slides on some take-aways that you wanna get O.S.F.-- e-- about, like, what we've been learning. So I can jump into those right now. But I didn't wanna just ramble at everyone.

ELIZABETH EAGEN:

Let's take a couple minutes for questions and then go back to the slides. (UNINTEL PHRASE).

MALE VOICE:

Well-- so I was wondering-- you-- you know, you had-- a kind of mix of different examples. Some where the data was sort of highly structured from the government. Had been con-- collected and organized. And you could apply, you know, machine learning or-- or whatever to pull-- things from it.

But then I think I were some other examples where it was sort of like a pile of documents. Or, like, a lot of P.D.F.s, you know. And I'm wondering from our perspectives, I think everyone, you know, interacts with probably more with the pile of documents, less with the government data. So-- just wondering on your end what are the sort of different outcomes? Or are there that, you know, that come from that.

JAKE PORWAY:

Yeah. It's-- it's a great question. And largely-- the nice-- the nice thing is, the dark secret is that whether it's non-profit or for-profit, people are mostly in the pile of documents phase. So rest assured it's okay. (LAUGH) It's like data therapy.

But I think-- so there are a couple things to say about that. One is that there's almost, like-- and you can almost always do something with that data. So that's what's been-- really excited our projects. We had-- the D.C. Action for Kids example. Yeah, some of that-- some of the well-being data came from the government. But a lot of it was still lot-- like, just scans of P.D.F.s.

Came in ill format and Excel files. There was a ton of data cleaning involved. But there were some-- a nugget of goodness in there. So I would say that, you know, when we work with groups, we always go through the data cleaning process. Let's say in the-- in the sort of worst case where you can't do anything beyond that, that's still an amazing learning lesson. I mean, we've had groups who they have, like, fancy database systems. They're using Sales Force.

And yet the way they were collecting data wasn't sort of well-formed. But just by going through that process of trying to answer the question, they learned what they would need to do to format it, right? Or the group that just sent us a bunch of Excel sheets because, you know, they filter out from different clinics and someone writes an Excel.

They sort of realized what it takes. You know, what they should do next time around. So they knew a lot to be learned there. And then lastly-- the Twitter example's maybe a good one. I-- I-- I mean, there's a big debate to be had about the value of Twitter data.

But I use it as an example because there is really high quality data out there a lot of the time that could supplement the pile of documents. So when we worked with the World Bank at one point and they wanted food price data. And their food price data is very sparse. They usually do it by visiting markets. But some-- data scientists found that if you just went to supermarket websites online and wrote some programs to scrape that data and organize it, then you would get a daily feed of price data.

So I would say even in that case where we walked in, they said, "Here's the pile of documents," they were able to say, "That's cool. Here's another data stream that with your guidance we could fill in the gaps of it." Maybe a longer answer but I think there's-- the-- the-- the take-away is there's always-- almost always something to do.

ELIZABETH EAGEN:

So let me kinda follow up to that, actually. And then I'll open it back up. How do you talk to you? (LAUGHTER) Because I think--

JAKE PORWAY:

That's a deep question.

ELIZABETH EAGEN:

--I (LAUGHTER)-- it is. I mean, I think what happens a lot when-- when groups that work for civil society causes find a resource like DataKind, they don't-- the-- the initial conversations are very clumsy. They don't know what it is they're asking for and you're not sure what it is they do.

JAKE PORWAY:

Yeah.

ELIZABETH EAGEN:

What's-- what's key to bring to that very first meeting? Or what's key to know about your own information before you sit in a room with a technologist or a data scientist?

JAKE PORWAY:

That's an awesome question. So you're exactly right that almost every conversation we have starts with a question. Starts with a question that never ultimately (UNINTEL) question. Because it is clunky. Like, what is data science?

I-- it's still a debate in the data science field. How would you even start expecting people to be able to know how to interact and ask for that? So the way that it-- works at DataKind is that it is necessarily clumsy to begin with. But the best thing to start with is actually about the questions your organization has.

A lotta people still think that data is the answer. A lotta people-- one of the most common requests we get is, "I've got a lotta data. Now what?" (LAUGHTER) Right? It's-- you know, and it's a (UNINTEL PHRASE) 'cause you have this-- it's-- it's-- it's opaque and it's this mystery.

There's gotta be something good in there. But of course like any good design process you need to know where you're going first. And so we always pull people back to, "What are your questions? What are your pain points?" In fact, one of the most ill-- illuminating questions we always ask is, "What sucks most about your job?" Because almost always that's where the-- the data science problems start to come out. So for example, a group came to us and said-- I'll-- I'll take the Red Cross. They said, "We've got all this data about where fires happen in Chicago.

"And so we want to transfer it to a new database system so we can (COUGH) visualize it and show people." And we said, "That's cool. But, like, what's your-- you know, what sucks most about your job?" And they're, like, "Oh, you know, what sucks is we-- we notice that people could prevent fires all the time.

"But we have-- and it's really simple. A lot of people still put space heaters under blankets. And that's, like, a real no-no. That causes a lotta fires and a lotta deaths. And so we do a campaign where we go out and tell people how not to do that. But you know, like, we're-- it-- it's-- it's hard.

"Like, we just have to go put out flyers. We don't have a lotta resources to do this." And so it was from there that we were able to say, "Great. So what would we need to target people most at risk of those cases? How do we do that most efficiently? Oh, historical fire data and data from the city."

And-- and from there we were able to work backwards to saying w-- where their data could play in. And where other data could come in to solve that problem. And in the end, they ended up with this targeted prediction algorithm saying these are the areas you need to go put your resources in.

That was way better than that database ever would've been. So think about the questions. What-- what's the most painful thing. We'll-- we're the-- we're the data experts, you're the subject matter experts. You come tell us what sucks most, we'll come tell you where data might be able to help.

And-- and on a practical level, if you know-- (SNEEZING) the more you know about your data, the more you can talk to the I.T. person and say where it is or what format it's in, the better it is. The way we usually do this is we talk to somebody, understand the-- the high level of the problems. Get a copy of the data.

Dig in and just, like, okay, regardless of the-- you know, we're not gonna be able to answer the questions about your subject space. But we can imagine what could be

possible. And then come back together and say, "Here's what we think could be done. Here's what you know, you know, is valuable. Where do they line up and what should we do?" Probably longer answer than you wanted but--

ELIZABETH EAGEN:

Long answers are good. This is the Open Society Foundations. (LAUGHTER) Over in the corner.

FEMALE VOICE:

Yeah, I have many questions. (LAUGHTER) I'm coming from working with the Accountability and Monitoring Health Initiative in the Public Health program and with the Global Health Financing Initiative. So many questions about the connection between data and social change, which is the title of the session.

But just from your perspective-- if you could give examples or stories of how you feel the data's actually being used for change. Because I think there's a difference between just informing or illustrating something using data. And actually that connection between what causes change and how much that evidence for that data is useful for causing change. And just how you have interfaced with that-- issue.

And then also I think a lot of us are in the space, especially with all of this open government and access to information and transparency-- movement that's happening. Where governments are increasing-- are pressured to increase access to information. But civil society doesn't always know what information would be useful and-- and-- and in what format. And to what extent are you helping people to know what to ask for?

JAKE PORWAY:

Awesome questions. All right, I'll start-- with the first one on-- on-- how-- how change is actually happening. I-- I think of an example-- I think it's called Climate Corp. Does everyone know Climate Corp and can help me tell the story better?

So the-- the story that I heard about these guys was they were helping come up with insurance for-- rural farmers. And the problem, of course, is that-- if you live in an information poor area you don't necessarily know what prices are going to be. You don't know what weather is coming.

And so it's very hard to work as a subsistence farmer. And so what these-- what this company-- this organization started doing was incorporating tons of feeds about weather data, about pricing data, to provide-- in-- insurance for subsistence farmers.

To say, "Hey, you know, we know that there's an 80% chance of rain that's gonna rain out the market. So it's not gonna be-- you know, people aren't gonna show up today.

So we can adjust-- insurance for you so you'll still get paid." And so this was a moment, like, you know, that was to me a great program.

Because it allowed them to use all these data streams to help provide a service that was targeted, that allowed people to actually, you know, do subsistence farming better. And one of the things that was so cool is they got so good at prediction that you actually-- you don't have to file a claim. They know enough about what happens with the weather in town that you actually just get a check that shows up to reimburse you for presumed losses because, you know, a thunderstorm hit or because a crop was dried out, ahead of time.

Now, I don't know if that's exactly the social change you're talking about (UNINTEL). That isn't even more, like-- no, that's a definite no. Okay. (LAUGHTER) It's a cool story, though. You can tell 'em that. (LAUGHTER) I guess-- I mean, this is-- I-- I-- I'd have to think about it a little more. I mean, maybe what you mean is-- do you mean, like, sustained, like--

FEMALE VOICE:

No, it's also that it takes decisions from powerful people in order to make change. And to what extent are you finding that there is or is not a connection between some of the data and the presentation of data? And the decision making that needs to happen to address an issue? I guess in our work we've found that there's multiple steps in that causal change between, like, information and-- and good data. And-- and actual change. And so I just wonder what you're learning about that.

JAKE PORWAY:

Oh, that's a great question. Okay, that's-- that's awesome. And actually it's one of the reasons I'm so excited to be working with O.S.F. because our work tends to end at the end of the project. So for ex-- let's say with D.C. Action, when you give them the map.

Great. You got the map. Now what, right? And we've been trying to work with partners to think through what would the rest of that change chain look like? And so what we've been learning about it is the-- I mean, there-- sort of-- this might seem like obvious design warnings.

But mapping that out up front is really important. People get excited about the shiny end but we really try to force them to play out for us, okay, what's gonna happen next? And-- what-- which policy maker's going to use that? What bill or-- will they write? How would we make sure that happens afterwards?

So getting the organization to think about that up front is really important. The other is actually finding organizations who already are committed to that chain. And I say that because we're finding a lot of organizations up front really want these results to impress funders. Really want these results because they're cool. Really

want these results because they somehow push their agenda.

And what we really look for are groups who are willing to hear hard truths. They're willing to say, like, "This may come up with something that is not favorable. And that's-- that's actually good thing. And how will that affect that chain?"

But obviously I'd almost-- I hate to dodge the question but I'll turn it back to you guys. I would love to hear how you've been-- how you create symptomatic change in the world. That's-- I can tell you where data might help. But I can't tell you the answer to all of that.

FEMALE VOICE:

Yeah.

JAKE PORWAY:

It's a start. And then-- oh, I blanked on your second question now. Oh, governments. Was it governments?

FEMALE VOICE:

No, it was just-- what do you ask for? How-- are-- are you helping civil society to know what to ask for?

JAKE PORWAY:

Oh, that was it. Yes, yes. (UNINTEL PHRASE). So-- that-- one of the-- okay. So the problem is-- there are a couple of problems here. One is that the government's opening up data without knowing themselves what the data is for. So that's like-- it's like throwing a bunch of wood-- like, random pieces of wood to architects, being, like, "Build stuff." I-- okay. But (LAUGH) I don't know what it should be for yet.

So that's-- that's one problem. It's not tailored to a purpose. The other is that people don't have-- data abilities in-house. So once you get that pile of wood, how would you even begin to assess whether it's good for what you needed.

So what we've been finding is that by bringing together these groups of the data holders. So, like, governments and social sector or civil society and the data scientists, that there becomes this really kind of, like, virtuous circle that happens. Which is great where the pattern is usually here's our question. We wanna-- improve maternal health. Great. Here's some data from the government about maternal health. Let's using the data scientist skills to try to solve it. Usually the answer is, "Oh, this-- we are missing this data. Or this data's not good for this reason. But we now know that we could ha-- you know, what we need."

And so that's been the most-- to me one of the biggest benefits is finding out what the data is that we need. And-- that the-- the social sector organization usually then understands what's possible and why they would need that data, which has been very helpful.

And we've actually-- you know, the government, to their credit, has come to us and often said, "We-- they're looking for that help as well. Where that we throw the data in the ether, we want people to use it. But we don't know if it's useful. It might be. You know, if you just tell us that we need one more column to suddenly, you know, add a public health issues. Like, that might be very easy for us to get. We just don't know if you need it." So I think that's really what we're hoping to accomplish is starting that-- (CLEARS THROAT) in that conversation, that dialogue with everybody. Yeah. That might be too-- (UNINTEL PHRASE) too big. But, you know, talking (UNINTEL PHRASE).

(FEMALE VOICE: UNINTEL)

ELIZABETH EAGEN:

Okay. (UNINTEL PHRASE).

FEMALE VOICE:

Hi. I just-- I-- thank you for your-- your talk. I really appreciate it. My question is about qualitative versus quantitative. Not necessarily meaning (UNINTEL PHRASE). But where does qualitative data fit in-- with what you do? Because I was particularly intrigued about one-- one of the points that was raised in the invitation to the talk saying why should we be wary about indices?

JAKE PORWAY:

(UNINTEL) worry about it?

FEMALE VOICE:

I'm-- I work in the field of education. And a lot of what we do is qualitative in nature. How do we know when (UNINTEL PHRASE) of this? How do we prepare teachers? Those kind of things that are difficult to reduce to numbers. And I just wanna know how do you--

JAKE PORWAY:

How do we work with that. Such a good question. (COUGH) There's a much bigger

conversation here around the idea of quantification in general. And I-- every time I-- I talk, I'll-- I'll just share this with you guys, too. I always mention this paper called *Commensuration As a Social Process*.

And I love this paper because it's all about the question of how do we-- or we as humans have a tendency to take our qualitative world and reduce it to numbers. But inherent in doing that, our natural social technological and political biases that not just are present but are so ingrained in us that we-- it's almost that we're unaware of them.

And so there's-- I'll-- I'll try to get back to your specific question. But I think this is-- a really important point to think about. Is that every time we're creating these numbers, we need to think very hard about how that number got there. And I didn't even just mean how is it digitized. But-- when you're looking at-- stop and frisk data, how many races are there? There's six. Like, why six? And, in fact, in some years it changes six to seven.

Right? So, like, if you just blindly loaded that data and started analyzing it and ba-- charts and graphs, like, you would be missing the fact that there are decisions being made in a policy level that influence your results-- that are very important. So this is a lar-- the-- that's very important for us to be aware of. And I think the more we get people talking about the provenance of data, the better the end result will be.

But to your specific question about qualitative. I think-- so to me qualitative data's super important. I mean, you can't quantify everything. Then we use it at DataKind all the time. When we work with these groups, we don't yet have, like-- a metric for how much they've improved their data literacy. We-- we're thinking about it or working on it. But we start with stories 'cause that's how people start to understand the world.

There's also-- so I think qualitative data is very important in that regard. I think there's also-- the very, like, techno-positivist view which is that quality-- qualitative data is quantitative data. So, for example, we had a group look at veterans-- coming through-- an improvement program for veterans. And they would write responses to surveys. And you may say, "Well, that's qualitative.

"They're just talkin' about how they're doing. And the program's not a number." But you could look at things. It was digital, right? So you can say how many words did they use? Are the types of words that they're using changing over time? Is their tone changing?

So this may not be what you were getting at, but we find a lotta times people who threw up their hands said, "Well, I just have stories." There's stuff you can learn from those stories. And, like, machines could help you understand about s-- the stories. So that's important, too.

So, I guess to-- to sum up that ram-- (LAUGH) that ramble. You know, there-- there's-- a real, like, process to converting qualitative to quantitative that we all sort of have to be aware of. It's inevitable. It's going to happen. And we need to be aware of how it happens so that we can question it.

But still find it useful where possible. And-- and that, yeah, qualitative data at-- if it's digital, there's something you can probably learn from it that you can use it-- can use data science on it. So I wouldn't think that that's not (UNINTEL PHRASE). So I don't know if that answered your question but--

FEMALE VOICE:

Kind of.

JAKE PORWAY:

Maybe not. (LAUGHTER) In a bigger conversa-- do you wanna-- I'm-- so do you wanna ask, like, a follow up?

FEMALE VOICE:

No. No, no, no. That's fine. (UNINTEL PHRASE).

JAKE PORWAY:

All right, cool.

FEMALE VOICE:

Well-- I'm trying to sum up all my questions into one question, which is basically when can we start? (LAUGHTER) Because--

JAKE PORWAY:

Yesterday.

FEMALE VOICE:

--I work on-- well, I work on ethnic profiling. And so what you just said really resonated with me because in different countries the people who are being ethnically profiled in Moscow are different from the people being ethnically profiled in Paris. Who are also somewhat different from the people being profiled in Madrid.

So-- you know, we're very aware of how-- when we're trying to track-- the odds ratio, how much likely is someone to be stopped if they're a member of an ethnic gro-- a particular ethnic group. Depending on the country, the groups being targeted are

different.

And have to be identified, you know, through some process of basically asking people, "Who do you think the police would stop?" Or, "Can you recognize that person as a, quote, minority person?" And so when we were doing this exercise in Moscow and everybody looked the same to me.

But the students who were monitoring it were, like, "Oh, that person is from the Caucasus." You know, they could tell. And by the end of it, I could tell. And then that became digitized data. But that's not my question. (LAUGHTER) Sorry. I'm too excited.

JAKE PORWAY:

Yes.

FEMALE VOICE:

No. What-- the particular question that-- we're working on right now-- has to do with documentation of identity. And the link to economic opportunities and well-being. So-- we all-- we work on-- act at-- access to nationality.

And even in countries where people formally have access to citizenship-- everything is so bureaucratized now people need I.D. cards to actually do anything. Even buy a SIM card, get a motor bike license-- certainly open a bank account. Being register as the owner of land that was never registered before.

Of course, voting. Of course, getting a passport to go abroad to work. You know, all this stuff is now linked to actually having documentation of your identity. So as part of O.S.F.'s work on the post-2015 sustainable development goals, we're trying to argue that documentation of identity should be incorporated as a part of development. You know, as, like, you know, economic opportunities are not available to people without I.D.s. So now I'll zoom in on Nepal. And we have several N.G.O.s that have been working for several years to help people get I.D. cards. And they've been collecting data like any old way.

So, we have 40,000 records in Excel of people from lower caste groups who never had I.D. cards before. And, you know, through the intervention of some program, you know, involving paralegals, 85% of them now have I.D. cards. So some people are saying this data's completely useless 'cause you have no comparative group.

These are only people who are helped by the N.G.O.. You know, and they're people from certain districts who we know are lower caste. And, you know, you don't know whether the same number of upper caste people would have the same problem.

You know, or the same number upper caste people could have their problem solved with paralegals. So this data's not useful at all. And so, you know, I just can't believe that that much data is really not useful. You know, and even if it is, we're about to

launch-- you know, some new work in Nepal because a lotta people, millions, still don't have I.D. cards.

And the question is, how do we help them get I.D. cards? You know, we still have a year or so left to run in which we could generate data which would still be useful in this conversation about whether documentation of identity affects economic well-being.

JAKE PORWAY:

Wow.

FEMALE VOICE:

So if we could start right now, (LAUGHTER) that would be great. If we could start tomorrow, you know.

(OVERTALK)

JAKE PORWAY:

Let's go. Yeah, no problem. So that's actually-- a really good point. And I-- I credit whoever was-- was crea-- was criticizing the sample, at least for being aware of statistical bias. And-- and sure, if you're tryin' to make a conclusion about everyone in the country, they're right.

You have a very biased sample. It's only people who came through this program. Then (UNINTEL PHRASE) I have sort of two responses to that. One is that you have to-- you know, it-- if you're tryin' to make a conclusion about all of, you know-- all of-- of Nepal, sure. This isn't good.

But if you're just looking at, you know, the specific group who are in your program and who might likely be in your program again, then there-- that's fair, I think, to-- to analyze to understand, you know, more about that group. You know, so a lotta times we'll have N.G.O.s come and say, you know, "I'm-- I'm using-- a cell phone program to get people on-- doing more public help."

But that's not representative of the country. Of the entire country. And that's fine because they're not tryin' to answer what's happening in the whole country. They're tryin' to answer-- they're like a business. They're, like, "What do I know about my customers?"

You know, like, or maybe-- you know, what can I learn from this to understand are there different segments? Are there people who, you know, have certain behaviors that would help me shape these programs when I'm working with that population again. So I think there's something (UNINTEL PHRASE). Other thing I would say is-- and this is a li-- this is gonna-- I'm gonna take off my statistician hat 'cause this is a

little bit blasphemous.

Which is that what business is starting to do as they've gotten more data is that they've stopped making causal conclusions. And started using correlations to drive business. There's a danger in doing that. That's basically saying, you know, two things are related to each other. As one changes the other changes. I don't know that that's gonna happen across the board.

I don't know that scientifically this one changes because this one changed. But if it's something as simple as when I send an email at midnight, more people click the links in it. Like, maybe it doesn't matter so much as long as I'm getting those link clicks.

You know, this is like-- this is very black magic-y. (LAUGHTER) It's not statistically sound. However, I do think there's a place for that in some of this work where, yeah, maybe your sample's not all of Nepal. But maybe there are correlations (COUGH) there that are valuable.

Maybe you learn there's a certain type of people who respond to this I.D. program really well. Does the matter exactly that scientifically in every hundred-- in every case that's the sort of reason that they're responding to your program? Maybe not, if it gets more people in the door and getting I.D.s. So, that's-- that's sort of my-- my feeling about it. Is, yes, there's-- there's something there. Or heck, at least we should look at it to find out if there is.

ELIZABETH EAGEN:

So I think we should take a pause there and go back to your slides, 'cause what I'm hearing is a couple of questions in the room about missing data. Some information about kind of the missing social graph. How do people fit together in ways that would inform the information that you find in raw data? And then I'm hearing a little bit about the policy decisions behind the-- the proven of data that you receive from the government.

JAKE PORWAY:

Yeah. I'd actually-- so, you guys probably have better and more specific questions than these learnings. So I'll go through these quickly just-- and we started to poke on some of them. But then I wanna talk (UNINTEL PHRASE) 'cause those sound really even better.

So really quickly, I wanted to highlight some of the learnings that we've done-- learned from workin' these projects with Elizabeth. And one that already came up is that translation is-- is critical. Like, this is actually one of the most important parts for us.

We thought we could-- our first model was to get data scientists and put them with N.G.O.s. And it turns out that, yep, pre-- they m-- little more babysat. Yeah.

There's-- I mean, I remember the first time-- I thought everything was goin' so well. I was listening on the check-in calls. Everyone was nodding. Yeah, yeah, yeah. And then they got to the end and some-- the data scientists stood up and they're, like, "Well, the P value of the correlation is (LAUGHTER)--" And I could just see everyone's face glaze over. And it's, like, "Dammit. What have we done?" So really our role is actually much less about data in-house now. That's what the volunteers do and, you know, we know. It's much more about making this easy, seamless. Make sure everyone's speaking the same language. That's actually become much more of our focus.

Then another thing is (LAUGH) actually, I used to think that data scientists only wanted to work on really cool big data problems. And it turns out that no matter what problem we've ever thrown out there, there's always someone who is so psyched to do it. And that shocked me because I wouldn't have felt that way.

We had-- one that came up was the Mayor's Office of Data Analytics in New York City came up and said, "Hey, we-- we take addresses and we map them to geographic locations in New York City. And this is how we distribute social services. If you get mapped to our system, you go to social service." Turns out Queens, for anyone who's familiar with Queens.

It's a nightmare for the system. Thirty-seven street app (?) five. It's, like, doesn't compute. (LAUGHTER) And that means if you don't-- actually, if you can't map that, they act-- those people don't receive services. So they said, we-- which is very tragic. Right?

So they said-- it's not that they don't but for this system they don't. So they-- so they said, "Would somebody, like, come and tune up this program?" I was, like, "That sounds so boring. Like, the data scientists that I know, they wanna look at cool data sets. And-- build network graphs."

I had 20 people sign up who drove into this for, like, 48 hours and were so excited about changing this. In the end, they actually improved. There were, like, something like 5% more-- Amer-- Americans. New Yorkers who were then mapped, where the mayor's office then came out and said, "There are now 150,000 people who can-- are eligible for services because of this."

So it was a really big impact. But I wouldn't have thought anyone would touch this. I would've thought just hire a consultancy to do it. So that's actually been really reassuring to me that people-- nerds like doin' nerdy stuff. (LAUGHTER) That's been really-- really helpful. And then the last one we sort of started to touch on, but is really important is that we talked about doing data science projects.

But these are, like, such a Trojan horse for us. Like, that's the exciting thing. Some people wanna do, it's a great starting point. But the success that we look at is actually not based on the project alone but on how much the organization now understands data science. Has now increased their data literacy. Like, the world we wanna live in is not one where we're, like, a consulting company.

Like, going around doing projects and then-- then you ask for more projects. I wanna see a world where, like, all non-profits have data science capacity in-house. Or-- or the very least know how to talk about it. Know when you need to ask for it. Right?

Like-- I think, you know, Elizabeth was the one who was saying, you know, everyone needs legal help. That doesn't mean you should all learn law. Right? So it's not like non-profit needs to learn to be a statistician. But you should at least know when you need to go talk to one.

So-- that's really been what we're trying to shape our programs to do is get people in the door, do something valuable for them for sure. But use that as a springboard to talk about how you think about this problem next time. How your data goes from being the-- the dump of files next time. So those were the three broad points that I wanted to bring up. But-- I think the specifics are actually maybe more interesting to talk about.

ELIZABETH EAGEN:

Okay. So go ahead.

FEMALE VOICE:

I have a quick question that a little bit unrelated to that. But I was curious about how you identify and understand sort of untapped sources. 'Cause I know that, for example, like, front line S.M.S. which is now being used for disaster response and (UNINTEL PHRASE) partner with Ushahidi.

And they're working with community health workers. But it was never intended to be-- I mean, from what I understand, a source for big data. And I'm just curious how you-- I mean, how do you go out there and say, "Okay, there's all this data and how can we use that as a source for," you know. I think that there's a lot of untapped.

JAKE PORWAY:

Yeah. Oh, you're absolutely right. I'm sure we're missing a lot of it. But-- a lot of it comes from-- the data scientists industry think about that a lot. Because once you start thinking-- once you start thinking that way of what is the data coming off of this product?

Suddenly it-- it-- it is like the-- the matrix. Or something like you suddenly see everywhere that data could be used that's coming off of something it-- that you wouldn't expect it to be. So a lotta that comes from our volunteers when they team up with an organization saying, "Oh, I know about-- this mobile mon-- money market that uses Impessa (PH). (CLEARS THROAT)

"I bet we could just ask them and they would give us the data and we could

understand exchange data." Or-- "Oh, I know about Oran (PH) released a bunch of cell phone data. It's anonymized but has a lot of location information. I bet we could look at refugee (UNINTEL PHRASE)." I'd love to-- clearly struck a (UNINTEL PHRASE).

So-- a lot of this comes from the experience of the-- the data scientists. But you're right, there are vast untapped resources. And-- the long term thing that I would love to see, that we've been talking about sort of-- building the future, is, like-- getting a data librarian. You know, like, and you go to the library. Like, that's what librarians are for.

You go to the library and you say, "I need this information about-- 18th century paintings." And they go find it for you. Right now there's nothing that's like that for data. You said, like, "I need, I don't know, wa-- water quality-- for the northeast." Like, I'm gonna go-- and that's a bummer. Because not only is it unregistered. Like, yeah, people are probably collecting that data for other purposes that's not even known about.

So it's a much bigger-- conversation. But I think a start would be having someone dedicated to going out and hunting it down and cataloguing it. It's kinda old school but I-- I think it would be hugely helpful.

ELIZABETH EAGEN:

Does anybody--
(OVERTALK)

JAKE PORWAY:

I have that would be the most (UNINTEL PHRASE) satisfying answers? Seriously, 'cause, like, every one I answered just kinda like-- (LAUGHTER) it's tough-- tough questions.

ELIZABETH EAGEN:

So, do we wanna talk about the ethics of data and the Hippocratic oath of data scientists? Or are there questions in the room before-- I'm-- I'm conscious that we have about ten minutes left of the scheduled time. We have the room for a while but ten minutes of the scheduled time. Anybody who has to leave might wanna ask a question.

MALE VOICE:

Sure. I know you do a lot on data analysis. Do you do any training on, like, data

collection? Particularly of impoverished communities or people that are in areas like slums and things like that where data scientists or data collectors are-- like, find it difficult to get to? And what kinds of technologies are available to them? What types of tools would they be able to use?

JAKE PORWAY:

So, we haven't done a lotta that yet. Most of our work has been around once there's some kind of data, how could we analyze it or learn from it? That's unless it's something like what sources are digital that we could use to supplement your data?

So like satellite imagery and things like that. Maybe a side of the conversation of bias there. So we haven't-- we haven't broken into that yet. But I think it's fascinating. And there are-- tryin' to think if-- it's-- does tactical tech do that kind of work? I feel like there were some this groups that are working on that, that I'm super impressed by. But that we haven't gotten into yet. But I'd love to hear if you think there's a good opportunity or what people are looking for (UNINTEL PHRASE). Oh, yeah, opened up.

FEMALE VOICE:

(UNINTEL PHRASE) school data (UNINTEL PHRASE).

JAKE PORWAY:

So, totally. All of those kinds.

ELIZABETH EAGEN:

Any other questions before we hit our ten minute mark? Hippocratic oath of data ethics.

JAKE PORWAY:

Yeah. I actually wanna turn that over to you guys a little bit. I'll-- I'll lay out the problem. And then I wanna-- I'm gonna be so curious to hear what you guys think or what you face. (COUGH) This system's really fun. It's great.

Like, you get data scientists who are coming out of industry who can work with non-profits on data the government's collected to hopefully form some kind of collective action. Or-- or better decision making. But the question then for us very personally that our data kinda work becomes this question of ethics.

So-- a case that came up that didn't happen with us, thankfully, was on a platform

like Ushahidi, which the woman who just left was maintaining-- that collects data about a-- community. If I'm a data analyst in San Francisco analyzing that data about a community that put their data onto a public resource where I'm gonna turn the results around to a publication that the government is going to use.

Like, who is ethically responsible to-- how-- who owns that data? Who owns those results? And what are my ethical responsibilities in any case in dealing with interpretation? So if I put out a report that eggs cause cancer because I analyzed the data on egg and cancer data-- is it-- and-- and somebody makes a bad conclusion about that.

Is it my fault because I didn't explain my methodology clearly enough and I didn't tell you how it should be used? Or it your fault because you're not data literate enough to be using it and you shouldn't be making those decisions? And that's something I think a lot about when we make these sta-- these kinda calls.

I don't know if that's what you were alluding to, Elizabeth. So guide me. And if you had something more specific or different you wanna talk about. But I'd be really curious to hear the ethical challenges that you guys are facing with data. And if you have any answers particularly.

ELIZABETH EAGEN:

And it doesn't have to be big and small. Just to add-- so--

JAKE PORWAY:

Oh, yeah.

ELIZABETH EAGEN:

For me the ethical questions that arise with data collection have been going on forever, before there were these advanced capabilities to do more with them, to publicize them, to shout them through a larger megaphone. The-- the question is.

Once you enter a system that collects your information about a human rights violation that's been perpetrated against you, who's responsible for maintaining the secrecy of that data? How anonymized can it really be? How does anonymizing phone records of locations actually get-- we can talk about that as well. I mean, I think, don't-- don't-- don't think of this as a-- problem of (COUGH) massive data or science-y kinda data.

It's really the ethics of using and handing over information about individuals who've been affected by human rights problems. Does anybody-- we could use more brainstorming on the pitfalls of this. Because if there is a question about developing a code of ethics for data scientists involved in civil society problems, then we should

address it. And we should think about it and know more about it. But we need to-- we need to sort of bring the proc-- problems to the surface to really figure out what that might be.

JAKE PORWAY:

Yeah. I have soap boxes. But I'm kinda curious to hear what other people think first.

MALE VOICE:

Does it depend on the data? So addresses in Queens, which are public, that's different from account of human rights violations. What about tin roofs? And-- what about tin roof data? Presumably the people would set-- I'm not sure (UNINTEL PHRASE).

MALE VOICE #2:

Yeah, right, 'cause people on that information, like, a house (UNINTEL PHRASE). You know, google images. Was it information about your house? And it was just a picture? Or when somebody interpreted it?

ELIZABETH EAGEN:

Right. Does the interpretation create more value in the data? And more information about you? And should we worry about that? Or is that inevitable?

FEMALE VOICE:

I don't have an answer to that question. But I think for me coming from being a researcher before coming into this world-- as you said, these questions have always plagued data. Whether it be qualitative or quantitative. I don't think we have-- the luxury as people who collect data or analyze it to feel like we're divorced from the outcome of that.

I think we do have a responsibility to be thoughtful about how it could be used. How it could be interpreted. Who is it going to? What does it mean for the people who are most affected by-- by the problems that we're studying? Does the interpretation resonate with the communities that you're trying to help? I mean, all of those things are questions that I don't think we have the luxury of not asking. So.

FEMALE VOICE:

But I-- what-- what I think might be happening is there's kind of an emergence of at least an awareness of a spectrum. So I think the way you presented it was more like binary. Like, is it your fault, is it my fault? But I think-- if we think of other analogies where there's been kind of an evolution of, like, an awareness of, like, ethical obligations of people engaged in certain space.

Like humanitarian actors, for example. If you look at the difference between I.C.R.C. and then M.S.F. Like, in the evolution of being aware of that ethical spectrum, I think that's kind of what we're muddling through right now with data. So I'm-- I'm waiting to see how that will map out. And then I think people will choose, you know, where they sit on the spectrum.

(MALE VOICE: UNINTEL)

JAKE PORWAY:

Yeah, it does seem like a lotta the debate right now is (CLEARS THROAT) digging out data. Data's bad. Shoved down. We can't give any data 'cause privacy can be violated. Or, yeah, it's inevitable. Just let it happen. Like, it's really grating to me that it's so binary. Because there should be a more nuanced conversation about what are the risks and trade-offs and certain things.

And I actually think-- so I-- I personally think that's really what I'm hoping that maybe you guys can-- we can all sort of help do together. Is, like, raise that conversation to a higher level. You know, what-- what bothers me is there being this information, you know, in asymmetry. So I may understand-- I-- I was using this example earlier of Chuck.

Like-- if Facebook asked me when I install an app. They say, "I wanna use your-- I want access to your friends. And I wanna know your location." And I know conceptually what I'm giving them. But I don't really know tangibly what that exchange is. Right? It's-- it's an-- I don't know the economics. So they could use that location data to know where I live. Know where I work.

Maybe just that information alone means that they can predict-- when I'm getting sick. You know. Because I'm staying more. And my insurance rate-- everyone always uses the insurance rates going up. (LAUGHTER) The most evil thing that can happen, right? And-- I-- I'm not saying that that's-- I'm not even saying that that's bad.

I just wanna know what I'm giving up. You know, I wanna be aware of that. If-- to make that decision, I want the full information. So one thing I think we should do more is raise people's sort of awareness about the-- about what their data could be used for.

The problem is-- the problem is, like, no one knows all of it yet. And who knows what new technology's gonna come out two years from now that when combined

with that innocuous data is gonna produce you-- had unintended consequences. So that's tricky.

There was another point I was gonna make. Now I'm forgetting. I think the other is merely that, like, I-- I do think anonymity is-- is pretty much dead. I mean, there-- and I-- maybe I'm okay with that. Like, there-- there-- I was reading a report that a-- chair, like, the new cars that are coming out are gonna have fi-- like, five pressure sensors in the seat.

So just based on the unique way that you sit in the seat will identify the driver so it can automatically adjust to you. If it's-- you know, you can set controls so if it's your 16-year-old kid, they can only drive during certain hours, won't turn on. Like, there's just-- we're not unique snowflakes. And there's so many recording things around.

That it's unavoidable. Like, it's gonna be so easy to figure out who people are. The question shouldn't be, how do we walk back from that? It should be-- well, how do we act in that world such that we preserve the values that we all care about? And how do we make that-- it's not an easy answer but I think it's more-- more productive than saying, like, don't. Like-- a do not sense list. Here's the-- how are we gonna do that?

ELIZABETH EAGEN:

So let's end there, actually, on that somewhat positive note.

JAKE PORWAY:

No. Positive note is that the world is full of opportunities. And there's more good that's gonna come from that data than bad.

ELIZABETH EAGEN:

Are there-- (LAUGHTER). Thanks, Jake. (APPLAUSE)

JAKE PORWAY:

Thank you.

* * *END OF TRANSCRIPT* * *