Jacquelyn: ... Very good news. Thanks so much for coming and thanks for attending Global Innovation Week and Dobility's session on transitioning from paper to digital data collection. My name is Jacquelyn Carlson and I'm a member of the Global Development Lab here at USAID.

[00:00:30] Just to get a sense of the audience here, how many of you have used mobile data collection tools before? All right, this is good. How many of you want to learn more about mobile data collection tools? All right, good news, you're in the right place.

[00:01:00] Just to give a brief intro about some of USAID's work and then I'm going to hand it over to Chris. Over the past couple of years, USAID has worked with governments and partners to encourage the use of digital data and technology where it makes sense, as I'm sure you'll hear more about today and also tomorrow. Switching to digital data collection has a number of benefits, having insight into where your workforce is, what they're doing and what is happening on the ground, helps organizations shift their resources in a much more timely manner.

[00:01:30] Evidence from case studies around the globe, proves that real-time data empowers leaders to make informed and timely decisions and can enable communities to experience those benefits. There's other examples of where digital data collection has had some real tangible impact, for me, I've been working in Libera over the past couple of years and ... Has anyone here who's been to Liberia before? Okay. Anyone who's been to Liberia? There, go. As many places around the world, the road network is quite a challenge so moving information from place A to place B

[00:02:00] can take weeks at a time, add that on top of rainy season, and you're talking months for a ministry of health to understand where their resources are and how to respond to new disease outbreaks. Because of this, the governments in that region, and also USAID, are looking at digital data tools to really make much more of an impact.

[00:02:30] Chris is going to talk a lot more about the platform and the company that's he's with today, but I just want to first give you just some resources to take a look at. I'll leave this up here just for a moment, and if you have any questions or want more information on this afterwards, please let me know. There's a number of people sitting in the back of the room today, if you want to raise your hand, Vivian and also Chris, they can provide more information on USAID's resources in mobile data collection.

[00:03:00] Just real quick, principles of digital development. We have some handouts so sitting up here if you want to take one before you head out. There's also a guide on paper to mobile data collection. It can be quite overwhelming to transition from paper to digital, and this gives you a step-by-step way to identify what your priorities are and so you don't necessarily have to reinvent the wheel every time, which those wheels can be very expensive. There's lots of different choices out there, there's lots of different tools, there's lots of different vendors and there's lots of different approaches, this can give you some more insight in how to think through that.

[00:03:30] That said, there's also different courses online. Tech Change is one provider of those. I highly encourage you to take a look and check it out. There's a USAID funded course on there that's free for anyone who really wants to try to experiment and test different tools and mobile data solutions.

Finally, there's a number of different resources on responsible data and real-time data for adaptive management, which means how can you use information to
[00:04:00] change what you're doing so that you're doing it faster, more efficiently, more inclusively. There's also a data action resource too that we have. I don't have the information up there, but again, please feel free to email us and we're happy to share more.

Like I said, there's lots of different tools out there, some are designed for low connectivity environments, some are not. Think through your needs carefully,
[00:04:30] experiment and most of all think through your user's needs and not just the people who are doing the surveys. Unless there's a very compelling reason, I encourage you to build off of the tools that people have also in their hands.

With that, thank you very much, and I'll turn this over to Chris from Dobility. Thank you.

Christopher: [00:05:00] Okay. Thanks, Jacquelyn for the introduction. I don't know how to shut this off so I'll just hand it off. Thanks everyone for coming. Just as a brief introduction, me name's Chris, I'm with Dobility, which most of you may not have heard of. We're a small social enterprise, we're five years old. Most people know us through our product SurveyCTO, which is a mobile data collection technology platform. It's used in about 150 countries now, thousands of teams use it for data collection. Great platform. It's not really what I'm here to promote today.

[00:05:30] What I'm here to promote today is thinking through this paper to digital transition, and in particular, by way of trying to establish my credibility here, this is what my surveys looked like, it would've been 2009. They came in these big rice sacks, they were really thick paper surveys, they took hours to administer. Our team would
[00:06:00] take filled out surveys, stuff them into suitcases and haul them on the Indian trains from where we were doing the data collection to the data entry facility that was in a different town and do the data collection. I'm not sure if this is going to work or not, but some of you mat recognize this if you've done data collection work yourselves.

[00:06:30] This was a video that one of my colleagues filed on their phone of our data entry facility that we've organized, and believe it or not this was actually a very well run data collection facility, or data entry facility, and we actually did a pretty great job keeping things organized, but you can see covers are ripping off the booklets as they move them around, there were just tremendous challenges, and I think in my
[00:07:00] own work I thought harder and harder over time about isn't there a better way to do this? And was motivated over time to transition to digital methods. This is all predating the social enterprise that I founded to help make those digital methods a bit more accessible. So I do understand the paper world.

[00:07:30] There are reasons why you might still use paper-based systems for particular areas. I think over time, the argument for using digital methods has grown stronger and stronger in more and more settings.

So talking a lot about digital data collection, I just want to start by establishing what I mean and some basic terminology. It sounds like a lot of you guys have already done mobile data collection so I won't dwell on what it means to collect data using
[00:08:00] mobile phones or tablets, but there's some different lingo that people use. In particular, there's this idea of computer-assisted personal interviewing. It sounds very fancy. It's an acronym, people like acronyms. It's essentially just talking to people and instead of writing down their responses on a clipboard or on a sheet of paper, keying it into a phone or a tablet. Some people use this technology for a
[00:08:30] wide range of applications, for example inspect health facilities, they might sit in a marketplace and observe, they might go visit small businesses and record prices. There are lots of different ways that you can collect data. A lot of it is based on actually interviewing people, but a lot of it is also based on observing, or inspecting, or doing other activities. Basically, I think what's common about mobile data collection applications is that you're using tablets or smart phones to actually
[00:09:00] record the data and you're filling out electronic forms.

When you have someone who's engaged in mobile data collection, they're using a platform like ours based on a particularly popular open source system, open data kit, then the actual tablet or phone interface looks kind of familiar and there's a fill-blank form button that people press and when they press that fill blank-form
[00:09:30] button, then a form comes up and they start filling that form out and they swipe their way through the form question to question.

You'll also hear people talk about mobile case management. Now, this is, I think, a terminology that grew up out of village health worker, community health worker, settings where you had health workers who were visiting the same women or the same households repeatedly, collecting information over time. Well, case
[00:10:00] management has grown to encompass a wide array of applications, but they actually look a lot like mobile data collection. It's just that instead of, for example, clicking a fill blank-form button, they click maybe a manage-cases button and when they click that button they get a list of cases that might have been assigned to that individual. Those cases might be water points, they might be schools, classrooms. Oxfam uses us in a wide range of refugee camp settings where cases are feedback
[00:10:30] from the community, so if somebody has a complaint or they have a question or a concern, it's recorded in this case management system and they can follow up with those people over time to resolve the problem and get back to people.

So what you mean by a case can differ dramatically from one setting to another, but at the end of the day you end up filling out a form on a digital device. To
[00:11:00] platforms like ours, it's essentially the same technology just being used in a slightly different way, with a different work flow behind it.

There are lots of other ways to think about digital data collection, so these are just

a few of the technologies, interactive voice response where you're either blasting out calls to people or they're calling in to say a former information system and they're pressing one for yes, two for no, we all use these systems and love them
[00:11:30] dearly. There are SMS systems of blasting out, whether it's some kind of appointment reminder or if you're collecting data you're actually asking people a question. I think this morning there was an example in the first session about are you sleeping under a mosquito net tonight? Yes or no? And you SMS back a response. Computer-assisted web interviewing, survey monkey call tricks, that kind
[00:12:00] of thing, that's digital data collection, telephone interviews that are being keyed into a computer or a mobile system ... Actually, 60 minutes is a really short period of time. I was told earlier that for this day 60 minutes is like an eternity, but we can't cover that much in 60 minutes so I'm not going to really talk about these methods of digital data collection.

Jacquelyn mentioned that there's a wide range of benefits to using digital methods,
[00:12:30] I'm also not going to talk about those for the most part. If you pin me down during the Q&A, I am going to try to wrap up as early as possible so that we can have a discussion and you guys can share what you've learned or ask questions. You can push me in this direction and I'm happy to talk about it, but I think that it is setting specific. The guide to transitioning from paper to digital methods has a lot of
[00:13:00] resources and a lot of ways to help you think about the trade-offs and when you would want to make this transition, so I just going to set that aside for the moment.

I am going to talk about something so I'm going to focus on the recent evolution in mobile data collection tools. In part, I hope a modest set of goals and I hope these
[00:13:30] are okay with you, what I want to do is walk you through how these tools have been evolving pretty quickly. These tools have become pretty professional, they've become pretty accessible and also quick and inexpensive. I'll show you what things looked like, and even if some of you had used mobile solutions early on in this ... By early on, I mean a few years ago. If you are early adopters, you might've used tools that were different, you might have hired consultants, you might have built custom
[00:14:00] solutions. There are all kinds of ways you might have engaged with this kind of technology and the thing is, the world has changed. If I convince you of anything today it's that things are different now and they're getting better all the time.

I'm also going to give you some of the basics that I think will help you think about some of the nuts and bolts of digital data collection and mobile data collection in
[00:14:30] particular and how you think about that. Originally, we had these really grandiose ideas about this hands-on workshop, and you guys all will programs like 23 forms by the time we were done with the session, then we learned ... Maybe there's not internet in this room and that made it a little more challenging, and so what I'm going to do is focus at a conceptual level, some useful concepts and facts about today's mobile data collection landscape, and then again, I'm going to push you to
[00:15:00] this online guide where you can read a lot more. This guide's actually a few years old, but I was pleasantly surprised to see that it was written from the perspective of "let's not talk too much about the specifics of what you can or can't do, or specific tools, but stay at this level of what are the trade-offs I'm trying to grapple with and how do I think through this process?" That stuff has actually stayed pretty relevant.

It looked as relevant today as a few years ago when it was created.

[00:15:30]

Recent evolution. Does anybody here use CS Pro? Okay. CS Pro has been the granddaddy of digital solutions generally designed for digitizing paper surveys, moving them from doing the data entry. So if you were running a data entry facility
[00:16:00] and you're trying to enter data from paper, you would use CS Pro. It's coming from a world, and using a methodology where essentially you're writing code, so if you are designing an instrument in CS Pro then you're essentially coding, programming an instrument and you've got lots and lots of people around the world who have the skills necessary to write this kind of code. In fact, they still use CS Pro and CS
[00:16:30] Pro now has an android app and you can use this for doing mobile data collection as opposed to doing data entry in a data entry facility. But rhe general idea was, "Hey, digital instruments are complicated. You need to basically have lots of if-then's and this kind of thing to control what shows up when and how to validate responses in that kind of thins." These are all real-world examples I got from somewhere.

[00:17:00] There was another software package, another early one later than CS Pro, but called Blaze, based on using laptops generally to collect data. Again, a scripting coding based may be a little bit easier to use. I know a lot of nonprogrammers who can write Blaze code, but still fundamentally it's writing code.

[00:17:30] Now, this open source platform I mentioned earlier, Open Data Kit, they started, and I should have a screenshot of it, but they started with XML and defining survey instruments. They started with an open standard for defining digital forms. It's called X-forms and it's based on XML. Still pretty technical, pretty complicated, not that many people in the world want to be coding surveys in XML directly in a text editor.

[00:18:00] But one of the spinoffs and contributors to this open source community said, "You know what? We're going to let people code their digital instruments in Excel in a simple format where essentially every row is a question or a calculation," think of it as a question, "and every column is going to be the type of the field. So if it's a multiple choice, or it's just asking for text, or an integer, or a note, and then what
[00:18:30] name do you want for the field, what label do you want to show up on the device. You can have multiple columns. If you want different translations, fine. You can have as many as you want. You just have another label column for every translation you want," all the way down to ... Here there's a little expression for a constraint column, which I'll talk a little more about, which is how do I validate entries to this field? They said, "We're going to take a pretty simple format that might be
[00:19:00] accessible to a broader range of people and then we're going to write the code to turn this into the ugly looking XML that the core tools want to be using for forms."

This is where I came into this process, and one of the first things I thought about is, "Well, when you make a typo in a field type, if I put in TXT instead of text, what's going to happen? It's not going to be really obvious? What can we do that's easy?"
[00:19:30] So we said, " Well, there's this conditional formatting stuff in Excel, so what if we make a template and we make worksheets for help, and we make the conditional

formatting so if you type a field type like note then it color codes and it recognizes that this is a field type and which columns are relevant for that. How do we take the next step?" This was a couple years ... We were pretty excited about this and then we created a library of sample forms and we're like, "Hey, can we just make
[00:20:00] this easier and easier?" A lot of people could do this. A lot of people know, whether you're an M&E Program Officer or you're somebody tasked with a health facility inspection program, or whatever else, Excel is in your wheelhouse.

So this worked pretty well, but at the end of the day somebody says, "hey, do this mobile data collection thing," like code the survey in Excel, we're like, "Well, I've
[00:20:30] got it in Word, how do I get into Excel?" They start with a blank Excel screen ... It's kind of intimidating. So by no means were we the only ones. In fact, we were late to the game in terms of saying, "Look, when you go into Survey Monkey, are you going to [inaudible 00:20:52] and you build a survey? It's not super complicated. You don't need to be confronted with a ... Why is it so complicated for international
[00:21:00] development settings and for mobile data collection? I want to work off-line, but that doesn't mean I want to have to use junkie tools and have things all more complicated," and so basically there was then another evolution, which is, and it's kind of hilarious how these things are layered because in our tool, and then in some others in this area, what happens is now we have this nice, more Survey Monkey like interface for designing your survey instrument and behind this there's
[00:21:30] still the Excel and behind the Excel there's still the XML.

It's kind of hilarious because when you go in now it's really easy, you click a big plus button and you say, okay, add my first question, and you put in, okay, well I want to be text, or I want it to be whatever, and we can integrate in lots of online help. So if you're not sure what this type of question is, you can click and learn more about it. We can do all this, but we can still basically turn this into an Excel.
[00:22:00]

For example, if you were on a plane or you were having to copy and paste between you have this big Word thing and you want to just get it into a digital instrument, you could still jump into Excel and mess around in Excel. Or you were in the field and you had no connection at all to the internet, you could still work on your survey and update ... You find a typo, or a translation that isn't quite right, and you
[00:22:30] happen to be in the field and you want to change it, you can still do everything off-line and in simple tools without having to be connected to the internet or anything like that.

So this kind of interface is now, I think, again .. The set of people who were willing to code things in XML, that was one set of people, Excel, that's another set of people, now we're reaching a point where it's becoming more and more accessible
[00:23:00] to a broader range of people and it's really easy to get started, and this isn't a SurveyCTO thing, this is a technology thing. This is just how this process works as technology matures and as something like this becomes more of a commodity.

It's not just building the forms. If you think about ... I'll circle back and talk about the key elements of any of these systems, it's not about building just the form, but
[00:23:30] even if you think about, for example, I come from the world of doing impact

evaluations and within impact evaluation if you're innovations for poverty action or poverty action lab, you've got annual training and you've got Ras and they study, they learn data and they've learned how to do code, for example, high-frequency checks, so that as data is coming in they're checking the statistical distributions to
[00:24:00] catch, for example, if this enumerators responses to this question are systematically different than this other enumerator. That's something you might want to look into. If there's a statistically significant difference in how responses are coming in from different enumerators or different teams, different regions, whatever.

So you've got people who are coding all this stuff up and they're learning best practices and they're following guides, and it's great, and they're really well
[00:24:30] resourced teams and they're super well-trained, but now what we want to do is we want to, again, take that cryptic complicated like process of writing code in Stata and bring it into an interface that more human program officers, M&E officers, can actually relate. So different platforms use different terminology in this respect so we've tried to boil down certain types of quality check that people want to build in
[00:25:00] as data is coming in to flag potential problems. People look for values that are too high or too low, outliers, but then this group mean is different, group distribution is different, I think this language we need to refine, so I think we think in terms of means and distributions, and I think that's something that probably in one or two versions of our product, you'll see that, that we use a different language that fits better with how people, program officer, might think about this.

[00:25:30]

But fundamentally, you want to be, for example, watching all kinds of questions and the responses coming in for statistical anomalies and for signs that there's something not quite right going on so that you can then investigate more deeply. The technology is systematically chipping away at how difficult that is to do, how much work it is, but then also what kind of person is capable of doing this kind of work.

[00:26:00] Same for data security. This is a funny thing, I still disagree with the bulk of the open source community in this, in the open source community they're like, "Oh, you can encrypt your data and you can treat it really safely, but you got to run this open SSL thing to create these keys, and you do that on the command line." I grew up with [Daws 00:26:24] and typing cryptic commands is fine for me. I can do that,
[00:26:30] but there was nobody else on my team in South India, where we were doing work, that this was fine for. Nobody wanted to type these open SSL commands and figure out how to create their encryption keys or anything else. But again, that was an easy problem to solve. You just need to humanize the interface for the stuff and say, "Okay, well, you want to create your own encryption keys."

It's actually not that simple because like we want it to be a simple web interface,
[00:27:00] but we never want to see the encryption keys as a provider, and so that's actually not that trivial. We do it in the browser, but we do it in such a way that it's all happening locally in JavaScript and in the browser so that when you're creating ... It says download private key file, but it's actually not downloading at all. It's actually generating the key locally in your browser and it's ... Nobody in the cloud, we're

[00:27:30] never seeing it, it's all happening purely locally but it's happening in the browser so it feels like Cloud, easy and whatever else. So it takes a minute to just say, "Okay, I'm going to name a key, I'm going to generate a new key pair, I'm going to save the private key and then the public key." That's it. It doesn't have to be that super complicated.

Speaker 3: What's a key?

Christopher: What's a key? Good question.

Speaker 3: [inaudible 00:27:54].

[00:28:00]

Christopher: Yeah. Fair enough. The techy-version is it's a public-private encryption key. It's not a password, it's like a password but it's superlong and it's super cryptic. If the NSA wanted to break your ... We're going to encrypt the data using this thing, and if the NSA wants to crack that, they're going to need a supercomputer working on it for like six months to be able to ... Or I don't know how long it'll take the NSA today.

[00:28:30] We're always surprised at how ... But basically the idea is that there's this big thing that only you have.

There's this cool technology where there's two different keys, there's a public one, and the public one is used to encrypt the data but you can't decrypt it using that, you need the private key to decrypt. It's a cool thing. We see the public key, and

[00:29:00] that gets actually put as part of your survey form, and everybody can have the public key, that's fine, and that's what they need to be able to encrypt the data. But then, in order to decrypt it, that's what you guard really closely. For example my original project in India, we had an institutional review board, we were collecting saliva and hair and personal financial data and all kinds of crazy things as part of an evaluation, and our institutional review board required that that data never be

[00:29:30] visible on a network connected computer, so we needed to use what's called a cold room computer. That's literally like sitting in a locked room with no network cables connected to it. Basically, you use this technology to make sure that ... That's the only computer that has this private key that can decrypt the data. Does that make sense?

[00:30:00] I just want to circle back briefly and talk about some of the basic concepts of how this works in practice. Let me just check the ... This monitor says it's 9:05 AM, which I strongly suspect is wrong.

Jacquelyn: It's [inaudible 00:30:17].

Christopher: Okay, okay. Great. There are a few key concepts, I've been talking about them already but I just wanted to circle back and just make sure that it made sense. This

[00:30:30] is a kind terminology that I think you'll encounter across different digital tools. But I talk a lot about forms and I often think I'll say survey forms but they're not always surveys, they could be an inspection checklist, a protocol, a form that you fill in to record prices at the market, it could be a lot of things, it's essentially a form. It

might be a survey but it might be something else.

[00:31:00] Remote sensing. This is kind of a cool thing nowadays. We're using satellite data to learn more about, for example, our areas where we're looking at the impact of rural micro-finance. Satellite data is great, but you have to train it, you have to ... This is ground truth in concept, you've got to collect data on the ground to link to the satellite data to actually make sense of what's going on, and for that, again, you need forms. In our case, we would send people out and they'd run around and
[00:31:30] they'd be like, "Okay, well, this is Patty. This is ground." We then use that data to train, to combine with the satellite data to train the algorithms for learning from the satellite data. So once we've told it enough times that this is ground and this is Patty and this is the building and whatever else, and so they're filling out forms for this. They're very short simple forms. They grab a GPS position, they might take a photo just for auditing, but then they'll record what's at this location, and that's it.
[00:32:00] That's like a form not at all survey, but again, using the same technology.

Different tools that'll look different, but I showed you the Excel earlier, as you're designing or interacting with a tool, you may be working in Excel if you're comfortable with that, you may work in an interface that's more drag-and-drop,
[00:32:30] what we tried to do is create an interface that parallels the Excel so that you could go back and forth so that there was an easy structured interface and you can drag-and-drop fields and organize them and work in an easy way for when you're getting started. But if you did want to drop down into Excel, you would be pleasantly surprised at how much sense it made because the language we were using ... Everything matches, the column headings in the Excel and everything, so we tried
[00:33:00] to ... Our interface is peculiar for that reason, but there are others that look a lot like Survey Monkey or a lot like other tools.

On the device, again, you might click fill blank-form and then basically you'd get a form. Usually the interface on the devices is pretty simplistic, and I think intentionally because in a lot of programs they use pretty low-end devices and also pretty old devices. I think my original project that grew into SurveyCTO in India, I
[00:33:30] think almost six years we've been using the same little galaxy phones. They're pretty scratched up, they're pretty old, but they still work and they perform well because the interface is simple. We're not trying to be really whizbang about what the interface is when people are actually filling out form.

Within a form, different people use different terminology, but often you'll talk
[00:34:00] about fields, and you talk about fields instead of questions because often ... Were any of you in the PPI presentation just before this? progress on poverty? Okay, so in lots of forms you will also want to make calculations. So in that, in the form for the PPI instrument, which is progress on poverty, it's asking a simple set of questions and then it's calculating an index and then it's feeding that index back to the user.

[00:34:30] There are lots of ways in which you actually want to be able to make calculations as part of a form and so that's why we call the rose in the Excel or the fields you add when you're designing your instrument in the web. We call them fields rather than

questions because sometimes you scan a barcode ... That's super cool, we've used that in a lot of settings in order to capture a unique ID really quickly, capturing a
[00:35:00] GPS position, making a calculation, taking a photo, that's a field, recording a video that's a field. These mobile devices are fantastic. It's nothing that we as a mobile data collection platform, we can't really take much credit for all the cool stuff that these devices can do. They can take photos, they can record video, they can record audio, we just get all that for free. So those are all you can think about all of those
[00:35:30] things as fields. Again, in the easy interface, when you add a new field ... This is in SurveyCTO so I guess this is what 12, 15 different types of fields ... It might be a date, an image, a GPS position, a multiple-choice, there are lots of options.

Another concept that's key is relevance. Now, this language isn't universal. It
[00:36:00] should be but it's not. This set of open source tools that are based on this one, very influential open-source platform, I think they might've started using this language of relevance, and the concept of relevance is just when should something appear? It might be a group of questions, it might be a single question, but the question you often want to ... One of the great, great, great advantages, so even though I said I wasn't going to talk about, but one of the amazing but subtle and seemingly
[00:36:30] unimportant advantages of a digital data collection is your instrument adapts automatically as part of ... If you're doing an interview or you're filling out a checklist, or anything else ... The thing is when you're designing the instrument, you constantly want be saying, when is this relevant? When should this show up? It might be only pregnant women in their second trimester, you want to ask this set of questions or you would only want to take a photo of the household, if the
[00:37:00] household had consented to the photo being taken. There are all sorts of ways in which you want the instrument itself to adapt to the setting and even to calculation.

If you calculate some kind of index, you might ask questions, you might randomize. A lot of people build in randomization because they're worried if they ask questions in one way or another, they might be leading and getting different responses. One
[00:37:30] way to know is to randomize and then see, do I get different data If I ask this way or that way? These are all things that you would control through this concept of relevance.

In the Excel, relevance is a column and it's an expression and it can be a complicated expression, but again, there can be a simple interface for designing. So I think in this case there's a little wizard and it's like, "If gender is equal to F and age
[00:38:00] is greater than 10, then we're going to say that this group of questions is relevant." So the tools have made this kind of thing easier over time.

Another key concept is constraint, which is what are the responses that I should consider valid? Sometimes people talk about field validation, that's the same idea, which is again, if I'm asking questions, when should I let them continue or when
[00:38:30] should I stop them and say no, -3 is not a valid age. 333 is not a valid age." The time to recognize that you've hit the three button an extra time is not six months later when someone's looking at the data and they're like, "That's weird. There's somebody age 333 that we talk to." The time to do that is in the moment, when

[00:39:00] they're talking to someone and they meant to type in 33, they hit the three button another time, they try to move forward and it says ... You would say, well the valid range is 18 to 120, say, and if it's outside this range I want to stop them and insist that they do something different.

[00:39:30] I'm going to speed by a handful of things and I want to get and have plenty of time for Q&A. I'll just mention a few of the more sort of advanced concepts that are a bit ... Some of them are a bit newer, some of them aren't covered by different platforms.

[00:40:00] The idea of audio auditing in some settings has been really revolutionary, which is all of these devices have microphones and there was a [inaudible 00:39:42], it was a meeting three or four years ago, where he asked me, he's like, "Can't you just turn on the microphone?" I thought about it and I'm like, "I guess we could." So then a week or two later, we had introduced audio auditing where you could randomize and record little audio clips of interviews, so particularly when people are doing interviews ... Now, again, you need informed consent, you need to control the data very carefully. It doesn't work in all settings, but in many settings, particularly when local firms are being outsourced, it's been an amazing way to corroborate the quality of the data collection, how people are asking questions, are they even asking the questions? That sort of thing. So audio audits are, I think, [00:40:30] something that is becoming more and more common.

[00:41:00] Something we call speed limits is people come up with different names, but again, based on the behavior, if people are moving through ... If you have a health facility inspection checklist and people are moving through it too quickly, you might be a little suspicious. So when we create fields in a survey instrument you can just say how long you think it should take and you can then say, "Okay I'll let people have a pass on three speed limit violations," but if they're systematically moving through this thing, faster than what seems possible, I want to flag it, I might turn on the microphone to listen in to audio records so that later I have something to listen to say, "Were these people just talking really fast or were they in a café filling out the instrument?"

[00:41:30] Published data. Jacquelyn mentioned real-time data, this is becoming a really big thing. People want more information quicker all the time, and so this has been an area of particular innovation and movement over just the last year or two. I would guess a year from now, the options for routing incoming data in real-time to different systems, dashboards, this is just getting better and better all the time.

Jacquelyn: Does that mean it's uploaded to the system as you [inaudible 00:41:55] at that moment?

[00:42:00]

Christopher: Yeah, that's a good question. In most settings, people aren't connected in real time when they're collecting the data, but so for example, in my settings in India, in villages, they would almost never have a 2G Internet connection, but every night they would sleep in a town, and the town would always have at least one provider,

and so they had a bunch of Sim cards and so they'd put in a Sim card. So they would try every night to basically sync what data they collected and so we would get it,

[00:42:30] basically, nightly, but it depends. Some projects, they go out really remote and it might be weekly, but basically as soon as the data comes feeding in, you can start routing that to places.

Google sheets is a great place. I could go on for two hours on data security and the ways that this can be safe and the way it's not, but I won't mention that here. But we have a great blog post on using Google sheets as a really inexpensive way to get real-time dashboards into a program that has basically zero budget, no tableau, no

[00:43:00] consultants, nothing else, just scrappiness. You can use Google sheets and you can actually do some pretty cool things.

We've been working, and I think a lot of other people have been working on, just getting better and better about tools for monitoring data as it comes in and learning from it, and particularly understanding where there are potential

[00:43:30] problems. I'm just going to zoom through those things and open it up to questions.

Good. Usually it's hard to get that first-person to break the ice.

Speaker 4: Thanks. I was wondering if you have experience in doing research? I've heard it maybe referred to as usability testing but I've maybe heard that term used for

[00:44:00] different types of research, but around what ... Are you providing the right information to get people to answer, to respond to the questionnaire, or respond to your data collection? so if you're doing an SMS, do you say official and put something in the SMS to get people to respond to it, or make them know that

[00:44:30] that's the data collection to expect? Or if you're doing an email and it comes from the government do you put official, or urgent, or please respond kind of thing?

Christopher: Okay, that's a great question. I'm going to start with a disclaimer and say it's a little bit above my pay grade, only I'll put my academic researcher hat on. As a mobile data collection provider, our ambition is to empower the people who can figure that stuff out, make them not worry about the technology part and focus on that.

[00:45:00] But that kind of question is really hard to answer. Some organizations, so we work a lot with IPA J.Powell, organizations like these, and they'll do randomized tests to see ... So for example SMS surveys, they have an attrition rate generally of 70% to 80% in terms of very few people actually respond to those, and if they do respond they may not respond in the right way, and understanding what's the population of

[00:45:30] people who are responding and what am I learning about them? How are they interpreting this? As an economist, lots of economist don't pay enough attention to survey research and methodology. There's cognitive interviewing, there's ... It's humbling when you take a set of questions and you actually go and ask somebody

[00:46:00] what those questions mean to them and ask them to talk through how they're responding to those questions. We're often surprised, shocked, outraged that actually they're answering a totally different way than we expected, or they're not responding because they thought it was this and it was really that. It's a huge area of concern.

[00:46:30] Now, I guess my hope is that if people ... If you don't have techies building the instrument and if it's not this model where we have this centralized technical team and we send them an instrument and then they code it and then they send it back to you and then we deploy it in the field. That's a really heavy process that doesn't allow for tight feedback loops. It doesn't allow for a lot of learning. Our ambition, and I think what's happening, is that now the technology is becoming such that the people who are closer to the field can actually go out and pilot and learn and ask people questions and figure out what's working and actually change in real time
[00:47:00] and say, "Okay, this isn't working. Let me try this." Basically make that process easier, but it is really hard, and I think people do need to experiment a lot and talk to people a lot and test a lot, pilot a lot, everything they're doing to collect data because otherwise, yeah, it's not clear that data is super high quality.

[00:47:30]
Speaker 5: How are you dealing with translation? A lot of the enumerators don't speak English, the people you're interviewing don't speak English-

Christopher: A good American, "I think everybody should speak English."

Speaker 5: How do you deal with the translation? On the form it's easy because the question can be written, but when you want to translate this back to USAID or to anybody, they need to be in English. What are you guys doing?

[00:48:00]
Christopher: Yeah. The way, again, most of these tools, the way they've evolved is essentially pretty agnostic about the languages, and so essentially when you create a new form you decide what language you use and you use as many as you want. So in certain settings, whether if it's in Ethiopia or if in some settings you actually have one instrument even in one region and you have enumerators who are using different local languages or dialects, and so basically, generally, the platform, it just
[00:48:30] has ... You have this form and it's just got a set of translations for this language and that language, and on the device you choose and you could even go back and forth within it. So if you were administering the survey in a household, you might speak with this person in this dialect and then switch, and so you can do that pretty easily on the device.

In the backend, then you would generally ... One of the languages would be English,
[00:49:00] and so when you report out the data, you push out the data, all the labels, if it's multiple choice, is it yes, is it no, is it C, so that all ... You would say when you push the data out which language you want it rendered in, basically, what the labels are.

[00:49:30]
Speaker 6: Thank you for all this information. It's been very interesting. I'm a little bit of an evangelist for digital data collection in my organization and ... We're getting there, but one of my recurring nightmare is the question of, "Okay, that's great. This looks wonderful. Yeah, the data, it's rich and definitely better, but is it cheaper?" And answering that question of realistic we're working within a pretty tight budget and
[00:50:00] we hadn't planned for this, how can we show that it's worth it or just can you make

it work with this budget? I found that a lot of these platforms really are a little bit expensive when you're working with multiple collectors and you need all these users and it's recurring, do you think that ... I guess the pricing model within ... Admittedly, I don't know on the pricing model, but do you think that it is actually

[00:50:30] accessible enough for organizations that aren't primarily data collectors?

Christopher: I do, but I do acknowledge that that's something that's been evolving. In fact, we've had this debate recently about the ways we measure social impact as an organization, and one of the things that I've champion, which has been a little bit controversial, is that one of the things we've done is come into this market and just

[00:51:00] ruthlessly drive the prices down. We've put a couple of people out of business and others have gad to drop their prices, they've had to drop the per user nonsense, and I view that as a win. For us, in terms of revenue, if basically we beat somebody else up and they drop all their prices and then we're not seeing that revenue, but as a sector and as a ... I think we are achieving our social mission, and that's a little hard to measure, but that's been happening really pretty quickly, I think.

[00:51:30] There are some entrenched technologies that are used, that, for example, have relationships with donors that and are at the right conferences and they've been able to keep prices at a ... But there are scrappy new entrants all the time, chipping away at that, and so I do think that the pricing is going down.

The trade also, again this report, for us it was a no-brainer, in India the printing

[00:52:00] costs were actually surprisingly high so it actually ended up cheaper ... We didn't have to do a really complicated cost-benefit. It was cheaper to buy 30 of these little galaxy wi-phones and we recouped just in the printing costs we saved within three months or something, and then we continued to use that for the next five years, those devices. So it depends a bit on the setting and the economies of scale, how much data you're collecting, but I do think, again, even if you looked at it today and

[00:52:30] you said it was on the edge, three months from now it'll be cheaper and you should look at again.

Speaker 7: I can't help but wonder what happened to the women who are collecting the data, the paper data, are they helping you with the technology or what are they doing

[00:53:00] now? Because now they're out of work. Just out of curiosity.

Christopher: I actually think enumerators ... So I have a huge amount of respect for the people who actually go around and collect this data, it's actually really hard work. I've done it myself, I spent a lot of time accompanying them. Working with a digital instrument is so much easier because you can build into the screen interviewer instructions, you can build in this logic about what appears when. On our own paper, we had arrows everywhere, it was- What's that?

Jacquelyn: You're not carrying around giant packs.

[00:53:30]

Christopher: You're not carrying around giant packs of paper. It's cognitively super difficult to maintain, like if you're interviewing someone, maintain a connection with them and

try to figure out, "Well, am I skipping to 3B or 37C?" The digital tools have made their jobs so much easier, I think.

[00:54:00] Now, it has offered new ways to monitor them and a bit of that is Big Brother, and it means that if you were sitting under the proverbial tree and filling out survey forms and you were happy doing that with paper, that gig is up. Now you have to tromp out, even in the rainy season, and find that household and we're going to check the GPS position, so the job's gotten harder in that respect, but I don't know.

[00:54:30] There are other people trying to innovate ways to put the whole idea of people interviewing people, or going around in person, this whole crowdsourcing data collection, there are people who are acting as disruptors, who want to just get rid of this whole thing because it's expensive and complicated. That's going to put a lot of people out of work, I think eventually. But yeah, I think this technology has made improved quality of life for those people.

Speaker 8: I think we're out of time.

Christopher: [00:55:00] Are we? Okay. We're in the evidence resource hub. There are a couple of us, my colleague Ruthie and I are there ... We have five more minutes? Okay. False alarm. There we go. I think you had a tentative-

Speaker 9: Yeah, I just wanted to figure out what has been your biggest challenge working on this platform in India?

[00:55:30]
Christopher: Biggest challenge? God, there have been so many. I was talking to Jacquelyn a little bit before about that our ambition is that this technology is not sector specific, but the thing is, as people think ... Like the language you use, if you're in health, or you're in wash, or you're an ag economist, you think in some different terms and different examples make sense to you, and it's been a real challenge for us to think about, how do we make this generic technology but make it in a way that connects [00:56:00] with all of these different people doing work that shares so much more than not? But if you're an M health person and you come in and you sign up on our platform, you get a bunch of examples that actually were created for development economists, and it might not feel like the right fit, and so that's been a real challenge, and I think something we're continuing to grapple with.

You had a question also?

[00:56:30]
Speaker 10: Yeah. I guess my question is much more general, I'm thinking in terms of less of the one-off or the repeated evaluations, surveys, more like ongoing records, systems, thinking about trying to move countries towards e-health records. Do you have any thoughts about the general hurdles or challenges in that paper to digital transition?

[00:57:00]
Christopher: I think that that's linking data over time and managing data and updating data.

Those applications tend to get more and more complicated- Platforms like ours allow some of that, but it's particularly difficult to figure out the right way to secure that data so that it's secure but then you can update it when you need to and there's the right audit trail of who's changed what. There are a bunch of challenges out of that, particularly when it comes to health records. That, I think, are really
[00:57:30] challenging. There are organizations that are much more focused on solving those challenges, and I think there are lots of applications within the M health space that I think are essentially generic data collection case management challenges that can use generic low-cost solutions. When you get toward actual medical records and that kind of thing, I'm not sure that my logic holds up and so I do think having
[00:58:00] technology and having people who are embedded and spend a lot of time with the health ministry in Zambia to get the right system working for ... So I'm not sure how generic some of those things really can or will be.

No more questions? Yeah.

[00:58:30]
Speaker 11: Thanks. I'd be curios to learn what problem does this not solve, between paper to digital.

Christopher: You mean generally in terms of for program teams or for what does it not solve. I think it doesn't at all solve her first question, which is how do you ask a question in a way that gets you a meaningful answer, like what should you be measuring?
[00:59:00] Those kinds of things are a big deal and this technology ... The way I think it helps is it gets the technology out of your way so that maybe you can focus more attention on those things that matter, but it does not solve any of those problems for you.

I think in an ideal world, people would share more. We've been trying to promote more sharing, so when somebody does figure out, in this setting this way of asking
[00:59:30] this question, or SMS seems to work well versus not. Unfortunately, certainly in the academic impact evaluation world, there's not ... You would get a report about something that was learned in evaluation, but they would not report out on all of the things they learned, all their travails and actually trying to measure things well. Trying to figure out what are the right [inaudible 00:59:51], what are the right ways to share more of that learning rather than just this end result or this dashboard, or whatever, that's in some sense the end result, that in between there's a huge
[01:00:00] amount of learning and we haven't seen a lot of evidence that people are really good at sharing that level of learning. There actually doesn't even seem to be that much interest, honestly, in resource, evidence resource hub and ... Thanks again.