

Pulse on the Principles EP 2 Final Transcript

Claudine: We live in a time when the world is more interconnected than ever before. But this has never been more apparent than it was in early 2020 when the novel coronavirus, COVID-19, swept across the globe shutting down cities and overwhelming healthcare systems. While governments scrambled to respond to the crisis, the development and humanitarian community stepped up to offer their knowledge, tools, and decades of experience to help fight and virus. At the forefront of the response, digital technologies are being deployed to curb the spread. But how can we ensure that these tools are being used responsibly and designed in a way that includes rather than harms individuals? This is "Pulse on the Principles." Welcome to "Pulse on the Principles," a podcast series that gives you a live look at putting the principles for digital development into practice. I'm Claudine Lim, your host for our mini-season on the "Digital Principles and COVID-19."

In recent months, COVID-19 has massively impacted countless development programs and projects, with organizations scrambling to fulfill deliverables while also determining how to appropriately pivot their work to respond to the pandemic. As a result, the development sector has been overwhelmed by new resources and proposed solutions, many of which promise that its plethora of data will help curb the spread of COVID-19 and make for better decision-making. Before we get to our guests today, I want to remind you that for the latest news and resources, be sure to visit digitalprinciples.org and follow us on Twitter @digiprinciples that is @digiprinciples. You can also use #digitalprinciples. Don't forget to leave a five-star rating and subscribe wherever you get your podcasts for more episodes, which now brings us to our guests. To help understand how data can inform programming, but also the ways in which irresponsible data can have adverse unintended consequences, I'm joined today by one of my co-hosts, Laura.

Laura: Hello, yes. It's great to be here. I'm so excited.

Claudine: And another one of our Digital Impact Alliance colleague, Rachel Sibande.

Rachel: Thank you so much. I'm glad to have this conversation with you all today.

Claudine: Rachel is the program director of the data for development team at the Digital Impact Alliance. She's a technology expert, computer scientist, STEM educator, social entrepreneur, and innovator. Rachel is the founder and director

of Malawi's first Innovation Hub, an incubator for emerging startups, entrepreneurs, and innovators, also known as mHub. Apart from the hub, Rachel is chairperson at Girl Effect Malawi and board member at GiveDirectly Malawi. Since the establishment of the mHub, Rachel has championed the development and deployment of innovative technology solutions across fields such as elections monitoring, citizen engagement, and agriculture in Malawi, Zambia, Tanzania, Mozambique, and Zimbabwe.

Rachel has over a decade of industry experience spanning academia, development, and social enterprise domains. So, Rachel, let's just jump right in. There's been a lot of buzz about using data for development and mobile data, including its utility in the COVID response. What does mobile data look like exactly? And what are we talking about when people talk about using it to respond to COVID? And maybe can you give us an example?

Rachel: Mobile phones generate enormous amounts of data every day. As people transact, as people move around, as they make calls, as they send SMS and do all sorts of transactions in the mobile phone, there's a lot of data that is being collected on various mobile networks. This data includes call detail records, profile data, location data, and spend data to mention a few. For example, if you look at call details records, they will have a timestamp for when a call, or an SMS, or any other transaction was made, who it was made to, what time that transaction was made. And it also does correlate with the tower location to which the person making the transaction was connecting to. So, they are large volumes of data that are collected every day through mobile network operators.

And we have learned over the years that when such mobile network operator data, for example, in particular, call detail records are combined with other types of data, such as satellite data, health data in a safe and responsible way, they could produce valuable insights to help us deal with a range of development and humanitarian issues, including the COVID-19 pandemic.

Now, Claud, you did ask to know specifically how this sort of data can be used to respond to COVID-19. Now, I'd like to give an example, using mobile network operator data in a safe and responsible way, and when it's combined with other data sets, we can actually start to understand mobility patterns of people. And mobility patterns can help us to understand better a number of things around COVID-19. For example, mobility patterns can be used as inputs to models in modeling the risk of contamination of the population throughout the country. But governments might also be interested to understand hotspots of COVID-19. And using mobility patterns, governments can help to identify key

transmission or transit nodes of COVID-19, particularly areas that are high at risk. We do know, for example, that several European governments are currently using mobile network operator data to predict potential hotspots of COVID-19 infection.

Lastly, Claud, using mobile network operator data we can also understand mobility patterns, which when combined with health data, governments can better identify health facilities that are at high risk of being overwhelmed by patients and understand how to distribute medical resources such as test kits, beds, and medical stuff. In fact, we are currently working on a very similar use case in Malawi in partnership with our partners Coopersmith, who are working on developing analytics for the Malawi government to understand the health system capacity during COVID-19 and to predict where new infections are most likely to crop up.

Claudine: That's a lot of information. And I think that also comes with a lot of responsibility. And I think the digital principles are a great place for starting a development project responsibly. There's even one called address privacy and security, but there's more to running an ongoing responsible data program. So, Laura, you've done a lot of work in the past on responsible data use. My question is beyond privacy and security, what other concerns should we consider?

Laura: Responsible data is a concept that's evolved quite a bit in recent years as we've realized that as humanitarian aid and international development actors, even when we're just collecting information for monitoring and evaluation, like a simple spreadsheet of people's basic information and whether they did or did not participate in the program, that's still people's personal information. And we still have to manage it, keep track of it, we could make incorrect decisions or assumptions based on that data. And so, a number of organizations in the space among them the Engine Room, and Oxfam, the Office for the Coordination of Humanitarian Affairs, and UN Global Pulse, among many others, have now put together an excellent catalog of resources that you can look to help you understand how to responsibly manage data in your programs.

Managing data well and thinking about data ethically and responsibly in the context of big data programs and mobile data programs like the ones that Rachel's describing is a slightly different thing. And there we're evolving some very collaborative peer review models to look separately at the ethics of a program and the data science, the model behind it. So, it's really heartening to see the field come together to recognize that this stuff is difficult and technical,

that we're all trying our best, and that there is a great benefit to be had by working on this together.

A couple of years ago, I worked on a study with the Future of Privacy Forum and the Center for Democracy and Technology, which tried to look at responsible data frameworks in their own words and compare and contrast them. And it came up with about five or six different elements of responsible data. And you named address privacy and security, which of course, is one of the principles. In addition to that, we think about basic things around fair information and privacy practices, many of which are often addressed by national law. So, the requirement to minimize data, to delete it when you don't need it, and to make sure that it's correct and that people can request changes and copies of the record.

But there's more to it than that as well around thinking about the benefits and risk of using this. What are you going to stand to gain from doing this work? Does it outweigh the potential risks of doing it? How can you be transparent and accountable about your use of data? And then now pushing very much into ethics and data justice and thinking about nondiscrimination and the impact particularly on marginalized groups. And those are all elements of responsible data. And DIAL itself is working on improving our approach to responsible data across those elements in our own work.

Claudine: That actually reminds me of the two principles, understand the existing ecosystem and design with the user. Our work should always love to minimize harm and maximize benefits to stakeholders. And that includes embracing community involvement and ensuring that flagging unethical practices are punished. But Rachel, I want to hear about the national context. How does this vary from one country to another and how are governments regulating these difficult issues?

Rachel: In our experience working in the various countries, we have learned a number of things in how governments are regulating these difficult issues around enhancing data privacy. First of all, is to talk about GDPR, which is the overarching global instrument for regulating data protection and data privacy. Having said that, it also provides very good guidance on how to use personally identifiable information in a safe and responsible way by clearly defining anonymization, for example, and pseudo anonymization. Which basically are mechanisms that are used to ensure that no personally identifiable information can be traced back to a person.

Now, that is very fundamental in our work, because most of the data that we've been using is at the center of integrating mobile network operator data with other data sets such as health data, satellite data, or data from the National Statistic offices. So, having to anonymize that data within the environment of the mobile network operator or the environment of the data holder is one of the key requirements that we have seen these countries enforce in a way to ensure personal data is not used in a bad way.

Secondly, we've also come across countries where they do have data protection laws. And we also know that there are countries that are currently in the process of developing data protection laws, whereas there are others which have clearly no instrument for data protection. Regardless of where we are, one of the fundamental elements that we've learned and seen is the need to actually put in place safe guardrails that can ensure that data is used in a safe and responsible way, nonetheless.

I also want to mention that the anonymization of personally identifiable information is not the only practice in enhancing safe and responsible use of data at country level, but we also take it all through how the data is processed, how data is stored, and finally, how the analytics are visualized to ensure that there's adherence to safe and responsible data practices. We've also learned from other countries that they are very particular and specific that data cannot leave the geographic jurisdiction of those particular geopolitical areas. Whereas for others, they've been flexible to have the data processed over the cloud as long as there's adherence to safe and responsible use of the data.

I wanted to mention that for countries that do not have data protection laws in place, there've been ways that they have provided guidance on safe and responsible data use using existing varied instruments. For example, you'll find that probably they already have an Electronic Transactions Act, or they could have a cybersecurity law in place. And so, we've collected bits and pieces from these existing pieces of law to come up with guidance on how to use data in a safe and responsible way.

Lastly, I wanted to intimate that within our work, we've also undertaken to get clearance from ethical review boards. For example, in Malawi, we had the Ethical Review Board within the health sector approve the use of health data in combination with mobile network operator data for the public health use case. We've also undertaken to engage telecommunications regulatory authorities to provide guidance concurrence and approval to these sort of mobile for development work. Lastly, we also have undertaken risk-benefit assessments and conducted internal reviews of the models and we are currently exploring,

getting some of our work to undergo external review by external entities to provide validation and credibility of this work.

Claudine: So how does DIAL go about incorporating responsible data into its data use work? What are some of the challenges organizations face in making this aspiration into reality? And how should it be considered as we navigate COVID-19?

Laura: It's a really tough one for any organization. But for nonprofit development, humanitarian organizations, there are a number of things about it that I think make it particularly challenging. One is that it's a real cross-organizational thing. So, you have to understand information systems, even your own internal information systems, the law, technology itself and how it works. So, for example, if you're hosting data on a server, you have to have someone in the organization who knows what that means in terms of the actual physical location of that data in the cloud. Does it still exist somewhere? Where is that and how is it processed? How is it transferred? Doing all that analysis.

And I think also, as Rachel mentioned, the law is in flux here. The law is way behind the technology and the evolving business models that we're seeing. Not just in less developed countries, but also in, say, for example, the U.S. and the UK, we're seeing the regulation evolve rapidly. And so, if you happen to be in a position of running a program that also operates across jurisdictions, so where data is generated in one country, processed in another, and then maybe the analysis is reviewed in another country, that can get very complicated and there may be very few lawyers who can tell you what that means. So, it's really complicated.

It's also fairly high risk. I think there's a lot of reputational risk for organizations and also legal and financial risk involved in getting it wrong as we're seeing countries evolve their enforcement mechanisms. So, Rachel mentioned GDPR, the General Data Protection Regulation, which is the pan European data protection regime that controls how data is managed in the European Union. It's become a bit of a model for nascent ideas to replicate such a thing in places like in Africa. So a number of countries are coming together to discuss that. But their enforcement mechanisms include Information Commissioner's Office, for example, in the UK, which can levy enormous fines for transgression. So, we're also gonna start seeing those kinds of things happening.

So, it's really challenging. People have to come together across an organization to figure out what data do we have? What data are we bringing in? Do we want it? Where did it come from? How are we managing it? And I think as we've

evolved in international development, a lot of these systems have evolved ad hoc. I would say probably cross-government is exactly the same. We talked about siloed, the sectoral development on an earlier podcast, where you have different departments procuring different IT systems to handle different things. You might have data in all sorts of places and being able to understand that managing correctly is a real challenge.

At DIAL, when we've been looking at this, we started from our values, which is something I really suggest because you need to be able to have it not be a new thing that comes in and feels new and scary, but actually is part of you being you as your organization. So, Oxfam for example, who are one of the first organizations to have responsible data policy, are a rights-based organization, so they grounded the policy in human rights. Similarly, DIAL has its own organizational values. And we looked at those, as well as those of our host organization, the UN Foundation, and meaningful regimes that spoke to us as an organization and developed our own responsible data values based on that.

And we focus on really weighing risks and benefits. And we have a risk-benefit analysis protocol that we now go through for each program to see what is the potential positive impact of this program? And what are the risks? And how can we mitigate those risks? And do they stack up? Do they make it worth it, essentially? We also prioritize transparency and accountability and demystifying the technology. So, being really clear about what we're doing, sharing our learning, and making sure that we can be held to account as well.

We're in early days of this and sort of still figuring this out, but I think it's a really good start. And we are looking to be more cognizant of the data justice elements of the work that we're doing as well in future years. But it's been a real privilege to be on this journey with Rachel and with the DIAL team as we figure this out, and it's hard work but I think it leads you to look very directly at your program and your impact and ask yourself some tough questions, which I think has been really valuable.

Claudine: And then Rachel, can you just chime in and tell us about how Malawi is committed to these kind of regulations and how DIAL might be involved?

Rachel: So, one of the things we've learned as we're doing this mobile data for development work in Malawi, is that it actually stimulated the thinking within the Malawi government to get started on the journey of creating their own data protection law, which were not existing prior. Currently, they have it in draft form. And as DIAL, we are committed to supporting the government there in the drafting of this data protection law in a sort of different way by providing a

space for civil society organizations, academia, and citizens to engage and provide input and be part of this process of developing the data protection law. Because we realize that ultimately, this data is about the people and there has to be a way to incorporate the voices of the people into the law that regulates how their data is going to be utilized.

Claudine: So to truly weave in responsible data use in programming is a lot of work, there's so much to consider. What other gaps currently exist? And why do you think they're difficult to address? Laura, why don't we start with you?

Laura: A gap in our field is actually capacity around this for all of us. And I think that organizations find it difficult to discuss challenges around this openly with one another and with donors, particularly. Donors themselves may not have a firm policy on how they'd like to see responsible data handled. And then everyone's working on their own internal policy. And with governments also trying to figure this out sort of at speed, it's very hard to have open conversations about this. So, I think conversations like this one, which can help to normalize some of these issues and present them as just part of what we have to grapple with as we use technology to help provide people with help are really useful.

And I think the COVID-19 response has really followed in the footsteps of the Cambridge Analytica scandal that related to the use of Facebook data and kept the use of data top of mind for the public. And so, I think it is getting better and there is increasing general knowledge about some of these issues. But actually, the process of figuring out what the social contract should be with this kind of data that is generated by private citizens using a private service in the course of their daily lives and how that can be used for public policy decision-making. How we have those conversations, as a national society and as a global society is a pretty tough one.

Claudine: Rachel, what do you think?

Rachel: First of all, the business models to attract private sector data holders are still not fully standardized and explored. There is a need to create incentives for data holders. We've also noticed that there's a lack of data science expertise, both in the private sector data holders, the government, and in particular, I want to mention the national statistic offices. Because these are the ultimate authority on data use within most of the countries. We've also seen there's a lack of a national digital and responsible data use infrastructure, both in strategy but also in investments. This has made it very difficult to make the use of these data models for decision-making routine, to make them replicable and

institutionalized, which is the ultimate place we would want to see this work go. And it becomes very difficult to work on data for development projects if there is no government approach.

Lastly, I want to talk about data literacy. Like Laura mentioned, it is very important for us to ensure that citizens fully understand the value of why these data projects would serve public good. Also understanding elements of how personally identifiable information is treated in this sort of cases. But even going further to explore mechanisms of how citizens can opt into this sort of initiatives in some cases and even opt out. I mean, I could go on and on, but all I'm driving at is the importance to enhance data literacy and evangelism to people to understand what it means to do this sort of work. There's still a lot of gaps and a lack of understanding within that space.

I also want to close off by emphasizing the importance of having champions within this sort of work. Within the countries at different levels, you need political champions, you need technical champions. You need champions within civil society among the citizenry to ensure that the message around data use for development can become a ripple effect that people fully understand.

Claudine: So, we talked a lot about digital literacy and making sure people understand what data of theirs is being used, how it's stored, what importance it has to policymakers. Do you think that there is sufficient work being done right now to educate people on why their data is so valuable and how it's being used?

Rachel: I don't think there have been enough efforts to enhance the understanding of people on the why this data is being used, the how it is being used, and why it is important. I imagine that more efforts have been put towards doing the technology itself, running the algorithms and the models. And it's typically been interfacing with governments and policymakers in this sense, but not necessarily a lot has been done to engage with the actual citizens.

Laura: I think where there is knowledge, it needs a little more nuance sometimes. And I think what will be important is for us to continue to understand how this kind of data interplays with the reality of society as it functions. And particularly thinking intersectionally about the impact on marginalized people, particularly people who have more than one thing that causes them to be marginalized. So, the intersection of poverty, gender, race, disability, age, all these things can mean that the way they present in data is different. So, I think as we do that knowledge sharing capacity building, it's really important to maintain research and learning around what happens when we use data in these ways, what pictures emerge, and then what the interplay is

between them and the reality so that we understand the potential impacts of using this kind of data in these ways on those marginalized groups so that we can better mitigate risks and show that impacts are really positive.

Rachel: Laura is right. Mobile network operator data is not the silver bullet to solving decision-making challenges in using data. As you know, not all people have mobile phones and cell phone towers in themselves are not available across all the countries as we know. The mobile penetration rate in most of the developing countries is not at 100%. We're currently working in countries where the mobile penetration rate is even much lower than 50%.

And so, we just want to intimate that it becomes very important to look at other complementary data sources. For example, grade 3 data, or data from national statistical offices if you really want to fully understand dynamics around the population. And to also intimate the analytics or the predictive models are purely predictive, and not necessarily an equal to, let's say, the number of people or indeed to the actual population density, but should be considered as estimates. And with these limitations and caveats of the fact that mobile network operator data might not necessarily represent some of the most vulnerable people in society.

Claudine: So, it sounds like there's no one-size-fits-all for responsible data. Existing culture, context, and behaviors change the implications and ways in which data is used. However, if there is one, what is the one overarching consideration organizations should take when they ask themselves if they're using data responsibly?

Rachel: I believe if there is that one thing that organizations should consider if they're using data responsibly is to ask themselves whether it is maximizing the benefit for the common person or causing harm? If the benefits are to the maximum, well and good. But if there is potential harm to the common person, then probably it is not worth doing it.

Laura: Yeah, I couldn't agree more, Rachel. I think putting people at the center of this equation is really important. And thinking intersectionally as well, thinking about people who might have complicated lives that don't represent everybody's or who have particular needs is really important.

Claudine: So, just like the principle that's always designed with the user in mind first and put them at the forefront always.

Laura: Absolutely.

Claudine: Rachel and Laura, thank you so much for chatting with me today. It's been really enlightening to hear just how much responsibility there really is in responsible data use, especially now as the development and tech sector build more COVID-19 solutions.

Laura: Thanks for having us and thanks for joining, Rachel.

Rachel: Thank you so much, Claud and Laura, for having me. It was a pleasure.

Claudine: And to those of you listening, if you would like to give us feedback on this episode, or any other topic in our episode lineup, you can reach out to us at principalsadmin@digitalimpactalliance.org. You can also visit us anytime at digitalprinciples.org and follow us on Twitter [@digiprinciples](https://twitter.com/digiprinciples). That is [@digiprinciples](https://twitter.com/digiprinciples). You can also use [#digitalprinciples](https://twitter.com/digiprinciples). Thanks for listening. We'll see you next time. "Pulse on the Principles" is made possible by the Norwegian Agency for Development Cooperation and is produced by Claudine Lim, Ilana Nelson, and Abigail Shirley of the Digital Impact Alliance. Special thanks to Podcast Village for recording, sound mixing, and graphics. See you next time.

