Innovation Webinar #3: AI-Enabled Data & Methods to Close Demographic Gaps for Children on the Move

Transcript

Michael Silberman

Thank you all for making it today. Good morning. Good afternoon. Good evening. On behalf of the International Data Alliance for Children on the Move, I want to welcome you. As you may already know, IDAC is a global coalition working to improve data and statistics and support evidence based policy making for migrant and displaced children. My name is Michael Silberman. I'm here with Kyriaki Kalimeri, my colleague. We both represent the Frontier Data Network at UNICEF. And this webinar series on Innovative Technologies for Children on the Move is organized by IDAC's Innovation Working Group, which is a project co-led by Save the Children and UNICEF with a wonderful network of colleagues from different organizations around the world. Our goal for the working group, as well as these webinars, is to foster collaboration and action using AI and new technologies so we can surface new solutions, identify opportunities, and solve critical challenges for supporting children on the move. This is our third session in the series, bringing together practitioners like yourself, researchers, academics, innovators. If you haven't seen the last few sessions, here they are on the screen. We'll share links in the chat. We encourage you to check them out. One was on predicting conflict-induced displacement and the other on climate-induced displacement, broadly speaking. There's more than that. But we welcome your suggestions for additional topics and suggestions, your suggestions for additional topics for additional sessions in the future. But let's talk about today, why we're here. Our session today focuses on innovative data and methods to address demographic gaps for children on the move. So, we're going to look at how population modelling and innovative data sources are currently being used to estimate populations of children on the move. Our challenge, as many of you will hopefully agree, is that children are often statistically invisible in population and mobility data. In many cases, child estimates are coming indirectly from assumptions about age structure or households or adult mobility, not child mobility. And so, while some of these estimation approaches and methods are often necessary to get something, they also bring us important questions about uncertainty and bias and appropriate use, especially when it comes to policy or operational decisions. So, to address all this, first, we're going to hear from Andy Tatum, Dr. Andy Tatum, on the opportunities and limits of small area modelling, where these methods work well, where they fall short. Andy is a professor of spatial demography and epidemiology at the University of Southampton, where he's also the director of WorldPop, leading a group of 40 researchers and data scientists. So, we'll hear from Andy here in a second. We're gonna explore novel digital data and what they can tell us about children's mobility. So to talk to us about novel digital data, we'll hear from Dr Emer Cleary, a spatial epidemiologist and senior research fellow with WorldPop. She has a background in microbiology, biostatistics, and spatial epidemiology. She'll be joined by Dr Haiyun Liu, a senior researcher at the Southern Marine Science and Engineering Laboratory, Guangdong Laboratory in China, and also a visiting fellow at WorldPop. And her research examines how Haiyan's research examines how extreme weather events and adaptation measures shape urban resilience with a focus on evidence based risk evidence based risk reduction strategies. Excuse me. Finally, we'll round out the conversation. Our third section

here, we'll be talking about mapping forest displacement. What can we say about children in IDP and refugee estimates? So, Dr Alexei Noskov is a research fellow at WorldPop. His research focuses on spatial modelling of forest displacement with an emphasis on improving IDP and refugee estimates using integrated geospatial databases, relational modelling, and fuzzy string matching approaches. And with him will be Dr. Sarchil Qader, a senior research fellow at WorldPop. His research focuses on developing innovative geospatial tools to support the effective implementation of modern censuses and national surveys. So, we will then go to questions. I will ask you now to please start adding your questions using the questions tool, hopefully on the top of your screen while presenters are talking. One question that we're going to ask everyone here to be thinking about, what would it take to improve child-specific estimates through collaboration between data scientists and child-focused experts? So, with that, I will stop my sharing and turn it over to Andy to take it from here.

Andy Tatem

[05:20] Great. Thanks, Michael. And yeah, thanks everyone for joining. Yeah, I will share my screen. And first, hopefully that works. First thing is to say is I don't think any of us on this have solutions to everything here. I'm going to give an overview of our work on population estimation and mapping, which and which covers the presence and estimation of children. And then, yeah, some of my colleagues will talk more about the displaced populations and the mobility angle. But yeah, so first of all, yeah, WorldPop, we're an applied research group at the University of Southampton focused on exploring ways and our understanding of ability to map small area demographics and dynamics focused on population counts, the demographics and mobility patterns. I'm sure you know the value of having accurate, reliable, small area population data for all kinds of applications. And obviously, if we are missing certain groups, they are missed out of all of these kinds of applications. The challenge, of course, is that certain populations are missing consistently, whether it's data that we do have that is trustworthy, but is it coarse resolution and so difficult for making decisions at smaller area scales, or the population data is outdated, it's incomplete, or we are just missing certain populations through them being not counted or inaccurate methods for collecting that data. So, at WorldPop, what we try and do is build up really a picture of the landscape and the situation where we cannot go and enumerate everybody all the time but making use of geospatial data sets. So, this may be buildings mapped from satellite imagery. It may be processing those to identify which buildings may be residential or non-residential. It may be identifying patterns in those data to pick out different types of neighbourhoods. It may be integrating sources of data that come from estimation of building damage in conflict situations, bringing in different types of infrastructure data, images of the earth at night that can tell us something else about different types of neighbourhoods, and also about how those populations change and move over time. We'll hear about later. And of course, the idea here is to build up this picture of the landscape so that when we cannot go everywhere all the time, we can have some proxies for why we may see higher populations or different population structures in different areas compared to others. It tends to look more like this, maybe a stack of 100 different layers. And typically, you may be looking at population data like this coming from a census. What we're talking about today is producing these types of estimates at the grid square scale. Those can be a simple disaggregation of that previous data down to those grid squares, or it can be an independent estimation using the set of geospatial data sets. So, we're dividing up the area of interest into 100 meter grid cells, estimating the number of people and something about their age and sex structures. And of course, these are at these scales are highly uncertain, but the value of having them at this gridded level is to be able to support decision making, whether it's estimates aggregated to administrative units or health

units or within settlement extent or linking with other types of data like distance to health facilities to identify those that are left behind or in need of interventions. The way we do this is very broadly bringing together some form of population data that we trust, some form of data that identifies where populations are not through the mapping of settlements. And then within those settlements, some of those geospatial data sets come in to try and enable us to estimate why we may see more populations in one area than another. And this is where either spatial statistical or AI type models come in. to produce estimates that are age and sex structured, ideally with some kind of uncertainty metrics around those. So we do this firstly on a kind of global scale, and this is work that finished recently to assemble more than a million administrative unit-based counts from censuses, from Common operational data sets from the UN is of course broken down by age and sex so that we can extract information on our estimates from these types of data on children and do things like this, looking at mapping of the proportion of the population under 15 years of age in 2020. And we also see that kind of movement of those dots. Each one of those dots represents an administrative unit and we see the world actually moving away from a younger population to an older population. We implement the machine learning methods to disaggregate those to grid squares and that enables us to produce these estimates of things like children under one year of age, working with UNESCO to identify what are the school ages for the different countries across Africa to be able to estimate school age populations or children under 5, very important for health delivery. And yeah, these data sets are openly available but obviously keep in mind the limitations with bringing together data from censuses and surveys. The other major part of work we do is co-developing estimates with individual governments, UN agencies, And this enables us to better understand what the needs are, what data exists to meet those needs, what skills exist to adopt and co-develop research and training plans to ensure that those estimates are actually meeting the needs and using the best available data. So, over the past few years, these are some of the countries we've co-developed those estimates with and there is a set of new ones waiting in the wings. Just to give you an idea of what the type of shape these might take, so they can be inputs coming from rolling surveys that are in this case for Afghanistan, where the last census was 1979, but more recent estimates were produced in those pink areas. It can come from health survey data from the distribution of bed nets in those pink areas in Nigeria, or it can come from a set of recent household surveys, the listings from those in Cameroon. producing those modelled estimates and then uncertainty around those estimates in terms of being able to communicate the fact that in some areas we are much more uncertain than others of those estimates. This can also involve filling demographic data gaps. So maybe just administrative data sources at core scales could be available, as was the case at the start of the Ukraine conflict. And so, using that to disaggregate to smaller area scales. or the case where there's under coverage due to inability to reach certain areas due to conflict and insecurity, and Burkina Faso here, and using approaches to fill in those gaps. So, there's tools to explore all of those different data sets, training materials for anyone who's interested, and we can share these slides afterwards. Just to show how these have been used for supporting different types of efforts towards child populations, this can be, for instance, the use of gridded estimates to delineate new enumeration areas for planning of new census. So, Satya will talk about this in his presentation. It can be to produce new estimates where a census has been outdated and use that as a sample frame for new surveys. It can be used to estimate exposure of child populations to things like conflict, flooding as a basis for health information systems or the denominators for health metrics. And they can be used in the planning of child health and education interventions. So, for instance, working with the Sierra Leone Ministry of Education to identify where schools should be optimally placed or the NPHCDA in Nigeria, looking at areas of cold spots of vaccination coverage and feeding into

vaccination campaigns. So, does that mean that we've solved everything? Certainly not. And so, I think, yeah, this is what we'd like to start thinking about together today is that there are, and these data sets I've shown, where we're assembling census data, we know that even in the United States, the last census had child undercounts. So, these are a major problem in the rest of the world. need to think about those populations that we may be missing completely from nomadic populations to slum populations to the homeless. Often we are missing certain populations that definitely includes children on the move. And this is a situation that in some trends is going in the wrong direction in terms of data, investment in data disappearing and obviously some household surveys really being underfunded. We also have the challenges of quality of those data that are going in. So, this is a study we did a couple of years ago to look at mapping out the variations in quality. And we see these high proportions of age heaping, an indication of poor quality data and the data of children of, in this case, the missing or implausible height for age data for children. Again, hotspots of poorer data there that we need to consider. And then of course, when we go all the way down to children on the move, if we're thinking about total population, we may have far more data of far higher quality than if we're looking at just breaking down by children or by mobile populations or by mobile children populations, we are really lacking data. So, I think there's just to finish off and to point the way towards the rest of the presentations, there is opportunities here by bringing together data sets, making them more interoperable to be able to integrate and draw the strengths of some to overcome the weaknesses of others. Important part of that, of course, is documenting clearly so that people can understand the limitations and what was done and why data sets are being used in ways that they may not have been originally intended for. The co-development, co-production, I think is vital in terms of we want to see any of these new types of data actually be used from start to finish, engaging with those people who are actually going to use those data sets to build trust, to build local ownership, if they're going to switch from the standard methods of straight line projections from the last census to a new approach, then there's really need to be trust built up in those to be able to argue to the public and to senior politicians as to why new approaches are being used. AI have put in a question mark there. are obviously many ways we can consider that, I think, and wherever we go across the world, there are often technical barriers in the types of methods and GIS in computer coding that AI I think as a primary approach could be really valuable in terms of overcoming those some of those and we're seeing that already. And then of course this decline in investment in data. I think we all have perhaps a responsibility to highlight data as an important public infrastructure where that enables us to actually do more with less. If we know where these populations are we can target resources more efficiently. It's often easy for politicians to see data as a way, as an easy cut to budget. So I'm going to just point the way finally to the rest of the team to present. Within WorldPop, there are, yeah, areas where we think We're doing okay, the rest of the global research community is perhaps doing okay in terms of mapping population densities in those middle distributions, but there are populations in the remote rural, nomadic, displaced, where there are real research frontiers and gaps. And at the other end of the spectrum, these high-rise buildings, slum areas, rapid areas of rapid change in urban areas. As well as, of course, we're doing OK in terms of total population estimates. But as soon as we start to break things down into smaller sample sizes, smaller areas, those demographics and those population dynamics become a challenge to capture. So, with that, hopefully I'm setting up the rest of the team to give some examples of work they're working on. So, thank you.

Kyriaki Kalimeri

[19:07] Thank you so much, Andy. And I must say for us, it's an honour to have you here. Another scientist and practitioner, I would like to thank you for the work that you've been doing and the help that you have that you have given to the to the entire community. As you said, this is this is exactly the place of the community that brings together both demographers, social scientists, experts on children. So, I would like to ask you a very difficult question to, if you, where do you think that these experts can be useful to you to help you produce better, to inform the population models that you're building and to help you make a more finer contribution to the population estimates?

Andy Tatem

Yeah, it's it is a difficult question. I think in every project we've undertaken where we've actually seen a solution being built and seen actually being used, it's been a it's been an incredible multidisciplinary team and not just multidisciplinary scientists who sit in isolation in ivory towers, but to be able to engage with those who understand the data and have actually been collected the data, those who are going to actually use the data is so vital to have in the room when these types of approaches and outputs are being discussed. We can end up with something completely complex and impenetrable to the world otherwise, and it just sits on a shelf. So yeah, this great this kind of initiative to start cross-disciplinary, cross-organisation discussions to hopefully try and break down some of those barriers. So, thank you.

Eimear Cleary

[21:57] There we go. Thank you. Yeah, thanks for inviting us to speak today. I'm going to speak through some of the data sources that we use for measuring population dynamics and some of the applications and limitations of these data, particularly with looking at children on the move and conflict settings. So broadly speaking, there's two different types of data sources that we use for measuring population mobility. So traditional data sources like census and travel survey data and newer data sources like smartphone data and satellite remote imagery and mobile phone data or call data records. So, call data records are data that are routinely collected by telecom operators. So, they're relatively low cost and easy data to collect and they're aggregated and anonymized for privacy. But these data can be quite biased in terms of mobile phone ownership, and the biases can affect representatives or representativeness of age and genders. And they can also be biased in terms of mobility type. So mobile phone data, for example, can't capture cross-border migration and the movement of children is unclear using these data because demographic information is absent. And also, this type of data might be unreliable in conflict areas where mobile coverage might be limited. So, we also use smartphone and app data to estimate population mobility. So, some of the data sets that we use are Facebook data sets, data that are available through the Meta AI for Good portals and the Humanitarian Data Exchange portals that estimate population movement. As an example of the publicly available data that we use is the Facebook movement distribution data. So, this data gives an estimate of how far Facebook users travel from home on a given day in four different distance categories. So, the figure on the right shows population movements on different days of the week in Bangladesh and activity space map data. So, these data are available by request from Meta and show the average home location of users based on a three-week aggregate of Facebook data and then the location that they spend on average in a three-week period other than their home location. So, these data can give us an estimate of daily mobility flows where people are spending their time away from their home location. a given day in a three-week

period. But these data all have different biases and limitations, so the coverage of these data might be quite good or quite sparse depending on the area that we're looking at. And then, for example, mobile phone data can be biased in terms of travel patterns. So wealthier people and long distance travel might be overrepresented, which might skew mobility estimations. And child mobility in particular might be overrepresented in mobile phone data compared with travel survey data. And again, we don't have demographics to be able to break that down and understand child mobility in better detail. And again, metadata might be biased in terms of the age cohorts and gender that use meta apps compared with census data. And there's also newer types of data that we're exploring at the moment to look at population distribution. An example of this is the Google embedding data set. And so, this is data that's built from aggregated search trends and point of interest data and business data from Google. And that's learned through a graph neural network and grouped together based on data similarity and proximity in space. And this produces a high dimensional feature data set that might be able to estimate population distribution depending on the features in the data sets. So, we've been looking at the utility of these data compared with WorldPop data for estimating population distribution. And you can see from this figure here that some of the features perform better at estimating population distribution than other. And the features that perform best are the features that are learned from the Google Maps and business data, which obviously makes sense. But these data also have limitations. So, they produce static estimates of population distribution. So, their utility for measuring short-term mobility flows might be limited, for example, in crisis situations. And again, we don't have any demographic information in these data, but they might be useful for exploring more long-term changes in population distribution. So, in order to understand some of the biases in these different types of data, we're currently planning a piece of work to conduct a travel survey amongst people looking for seeking a malaria diagnosis at health centres in Zanzibar. So, we're planning on conducting a travel survey and also asking for consent to gather people's individual mobile phone data. This will commence hopefully April, May of this year. And we'll compare these data with aggregated smartphone data and OpenStreetMap data to try to understand some of the biases in each of these different types of data and how these different types of data perform in terms of and estimating mobility. And so, to kind of to wrap up, one of the ways of overcoming some of the limitations and biases in these data might be to use a combination of different data sources. So, for example, census and survey data that contain more demographic information and can give more information about child specific mobility and then cloud data records and social media data that might give better estimation of short-term mobility flows and there's also different modelling. and data weighting approaches that can be applied to these data to try to impute missing data and address some of the representativeness and biases of these data. And there's some links in the bottom of this slide here with more information about some of the modelling and weighting approaches that can be used. Thank you. It's Tina. Should we go to? Let's go to Haiyan.

Haiyun Liu

[29:22] OK. Hi everyone. Here is Haiyan. Today I would like to share how we combine the Geo AI method and large scale human mobility data to evaluate disaster response policies and then reflect on an important limitation that is especially relevant for child focus research. The partial invisible or child in this dataset. I will first briefly introduce two empirical applications from China. First, we evaluated multi-hazard early warning systems using mobility data derived from over 1.1 billion mobile devices covering over 77 percentage of the national population in China. We examined the behavioural response to more than 21 warning signals during 19 landfalling tropical cyclones. Mobility data allows us to observe how people adjust their movement in real-

time when a tropical cyclone approaches. We see a clear reduction in mobility driven by both directly hazard exposure and early warning information. Because warning types, timing, and spatial coverage were across the event, we were able to isolate their average effect. We find that the combined warnings across multi-other type and administrative levels produces the strongest behavioural response. Compared with city-level warning alone, combined warning reduced mobility by 52% and increased the avoided population exposure by 57%. In addition, each additional warning types significantly shortened post-disaster recovery time, with the recovery duration reduced by at least fifty-five percent. We then applied a causal inference framework to evaluate heat adaptation measures, including cooling centres, urban greyness, and heat warnings. After controlling for city-level socio-economic and environmental characteristics, we estimated that the causal effects of these measures are mobility behaviour during extreme heat. Our key finding is that adaptation benefits are highly context-dependent. Cooling centres and urban greenness show limited effect during daytime-only heatwaves. And the heat warnings generally reduce mobility, but in some vulnerable agricultural regions, extreme heat warnings are well associated with increased mobility, likely reflecting economic constraints and prevent behavioural adaptation. This highlights that adaptation policies are not universally effective. From a child and equality perspective, this matters greatly. Policies that appear successfully on average may fail or even produce unintended consequences in vulnerable communities. And despite the scale of our data set, child remain partially invisible. Child under six are largely absent from our mobility data because they are not primary mobile phone users. In China, school-aged child increasingly uses smart watches, which allow partial observation of their mobility. However, child under 6 are not observed. We often assume that a young child travel with caregivers, meaning adult mobility can serve as a behavioural proxy. While this assumption may be reasonable in many contexts, it remains unverified. And because mobility data are provided as aggregated indexes to protect privacy, we cannot directly observe whether children's response differs from adults during disasters. And when we consider low-income and data sports settings, these issues become even more critical. Our methodology framework is scalable. Our mobility data is the mobility data Emma just mentioned is increasingly available in many low-income countries. However, demographic gaps and special bias are often malpronounced. In rural and low-in density areas, mobility data are often frequently suppressed for privacy reasons leading to underrepresentation of already margined populations. And the WorldPop team have demonstrated that a BSN imputation method can effectively correct spatial bias by combining social media mobility data with census-based population estimates. This provides a promising pathways for addressing spatial inequality in data. And our next step will be demographic correction. If we can spatially recover missing spatial information, we should also explore modelling strategies to account for missing child population, particularly young children. Well, thank you.

Michael Silberman

Thanks so much, Hayan. We're going to now switch from going from novel digital data over to our mapping forced displacement final section. So here we have. Thank you, Sarchil. Take it away. We have your screen up. We can see it from this end.

Sarchil Qader

[35:06] Awesome. Thank you very much. So, in the next 5 minutes, I will also walk you through some of the work that we've been doing. around exploring AI-enabled geospatial modelling to map forcibly displacement and integrate them into the national statistics. So, children often remain invisible in displacement data because records usually lack precise locations, are too

aggregated, and cannot be combined across systems. As a result, their specific needs, vulnerabilities are frequently missed in humanitarian planning. So, in terms of the data challenge, I think some of my colleagues have already mentioned, but displacement data is fragmented, uses different definitions and rarely aligns across a system. Key areas like informal settlements and secondary movements are poorly captured and survey lag behind fast changing realities. This makes it hard to track dynamic mobility pattern effectively. This is a proposed sort of like AI machine learning based workflow that could be potentially be used for the segregation of forcibly displaced person, including children. So, in this workflow shows how AI machine learning can help us produce high resolution age and sex disaggregated displacement maps, including for children. We begin with the harmonization and name matching, which is quite crucial here. And the method can help us clean and align the displacement data set that come in different formats. Next, we use household survey to extract key population characteristics, such as age and sex. We then combine this with gridded population data to estimate the displacement to host community ratio at a granular level. Using a stacked set of geospatial layer, AI machine learning, the desegregation model can then predict where displaced population, children and adults are likely located even in areas with sparse information. The final output is a high resolution grid or reasonably moderate resolution grid, and a polygon map showing displacement pattern by age and sex. It's important that we can see that the ethics, we should have a proper ethic in place to ensure that model outputs do not harm vulnerable population, reinforcing responsible and safe use of AI in humanitarian setting. So here we present a proof of concept demonstrating the application of this method to map refugee population in Cameroon in collaboration with UNHCR. The approach combines a simple disaggregation method within refugee camps and a machine learning model. applied outside camps to estimate refugee distribution. The figure on the right illustrate the visual improvement achieved through fine scale refugee population mapping compared with conventional UNHCR reporting. While the two approach shows strong overall correlations, fine scale mapping capture detailed spatial variation rather than representing population as a single geographic points. This analysis drew on a UNHCR progress registration database. The methodology has a strong potential for expansion to other contexts, including countries in Latin America. Recently, the data set was also used to produce low resolution graded refugee population estimate across 25 sub-Saharan African countries, different methodology for different purposes. However, we should Keep in mind that the quality and level of detail in registration data vary considerably between countries, which affect the precision and reliability of resulting estimate. So, the work extended beyond producing high resolution refugee population estimate. The output were integrated into the displacement survey design conducted by UNHCR and Cameroon National Institute of Statistics. So, at WorldPop, we have developed innovative geospatial tools that are used to automatically create and create digital national sampling frame. So, using these tools and leveraging the high resolution model refugee population data, a customized national sampling frame was developed to support the 2024 displacement survey in Cameroon. So, this approach addresses a critical gap in many countries where existing sampling frames are not adequately representative for forest displacement survey. So, in inclusion, I wanted just to say that AI and geospatial modelling are accelerator, not replacement for official statistics, which they can help us revealing children who are historically unseen. Disaggregated, highly solution data turns fragmented information into actionable insight for policy and humanitarian planning. So, making the invisible visible ensure every child counts in survey programs and decision. Thank you very much. Alexey, for you. Thanks. Sorry, I'll go to you, Alexey.

Alexey Noskov

[40:51] So, I'm going to talk today about mapping of forced displacement and what can we say about children in IDP estimates. And it will be about the project recently implemented for Nigeria and the Democratic Republic of Congo covering 2022 to 2025. And so, this project was funded by GRID3 and GRID3 supports health focused decision making in Africa. through high resolution population and boundary data in partnership with governments and the Gates Foundation and field campaigns and micro planning the IDP data delivered as a high resolution spatial estimate extracted and harmonized from and otherwise fragmented and heterogeneous data landscapes and updated to reflect resent displacement dynamics. But displacement data are often noisy and inconsistently formatted across sources. Addressing this requires for the string matching and reconsolidation of administrative boundaries and that differ across data sets. Although the project was not specifically designed as a child-focused initiative, child disaggregated data are presented within the integrated database and can be extracted where available. The framework therefore provides a strong foundation for the future child focused DP analysis. So, and basically our solution we delivered for grid three. So, we designed software and output and output project data set, and We included several data sources for our analysis. So, for Nigeria, we used mainly IOM, DTM data from 2022 to 2025, OSGOF and Grid 3 admin boundaries. And for DRC, we used OCHA IDPs data from 2022 to 2025. And basically, we developed software which consists of consists, which is consisted of five main tools for importing IDP spreadsheet data, importing GIS data, spatial linking of GIS and IDP data and we developed country drivers for each country's specific instruction, how to build holistic database and output data. And regarding output data, we have 204 JS layers produced for Nigeria and 216 produced for DRC. Yes, so here you see some visualization based on our output data set. Upper row is for... DRC data, illustration from DRC data, and bottom row, Nigeria. So, from left to right, the spatial granularity is increasing. So, you see examples of regions, IDP or regions, and regions of IDPs, IDPs destinations. returnees and for Nigeria to write maps, recent changes in IDP's awards and in specific locations. But also, we can extract specific children data from such a database. So, I've prepared the map from specific rounds where we can see the number of children in specific Nigeria wards and disaggregation of age and sex children groups. But so definitely we need to invest more resources to extract systematic data from such databases. Yes, so that's I wanted to present today. Thank you very much.

Michael Silberman

[46:11] Thank you very much. All right. Kyriaki, I'll pass it to you to lead us through some of the some of the questions. Please, I'll continue to add your we have a few questions in here already. Please, others, you're invited to continue as we have a few minutes now to for some dialogue.

Kyriaki Kalimeri

Exactly. Please put in your questions. I will start by asking the question that has been placed here in the chat regarding indicators like diaper or other child specific items and whether these are being used in improving the estimates or other types of this sort. If someone wants to take this one.

Andy Tatem

[46:58] Yeah, I was just typing an answer. I'm not aware of that those kinds of data being used. I guess it depends a lot. We need to understand the local context of how many children use the diapers, how regularly, how and that data would like, I guess, likely come from the companies

that are producing them and distributing them and therefore what data are they using to get their demand. So, there's lots of Lots of questions I would have about the use of that type of data and understanding the biases, but if we could understand those and dig down this, yeah, it's potentially, I guess, depending on the context.

Kyriaki Kalimeri

Sure, thank you. A question that I had while listening to your presentations and knowing your work, I would like to ask To what extent are the population products that currently exist capturing the specific patterns of children? Or they are inheriting the patterns of the total population? Because as we were saying in the beginning, in many of the digital data that we're all using, children are not specifically captured there. They are captured via their parents or someone else. So, I would like to ask What do you think about that?

Andy Tatem

[48:35] I can start going on the, on the, I mean, on the, I've covered some of this, I guess, in our estimate gridded estimates that it, yeah, the kind of patchwork of quality of data varies hugely in terms of whether there is child specific data going into models or we are producing estimates at small scales and then and then using a much coarser scale estimate of the of there being a certain percentage of children in an area and in those cases the uncertainty is much, much higher than if we are making use of yeah household survey data where the specifically numbers of children are there. But yeah, I don't know, maybe him or Haiyan can talk about the actual digital data sets. the kind of mobile based ones.

Eimear Cleary

[49:31] Yeah, I mean, I think, so first of all, the data is anonymized and aggregated. So, in terms of real numbers, there's no real way of knowing and doesn't contain demographic information. So, there's comparisons that can be made with population estimates and census data to try to get an idea of, you know, the There's a comparative kind of idea of the Facebook users, for example, compared to enumerated populations in the census based on based on age and gender. But I guess it's context dependent as well because like so if in the in the. In the presentation that I gave on the mobile phone data, I referenced a travel survey that was conducted by a colleague of ours based in the States who that showed that mobile phone data like massively overestimated mobility and by children compared with actual data. But I guess. Mobile phone data itself might capture children's mobility to different extents depending on income level of the country and individuals. So yeah, it's difficult to quantify, but definitely getting more data from different contexts on actual mobility patterns amongst different demographics and then being able to compare it to digital data will give us a kind of a better insight into how representative they are.

Kyriaki Kalimeri

Thank you. And just because we're going towards the end, I would just like to ask you if there is something we can do to overcome the biggest issues we have. So, if we look five years ahead of now, what would success look like for children population data? What collaborations are we missing? What is it there that you would love to have and how we can contribute as a community to that?

Andy Tatem

[51:39] I can start that. I think often when we are starting a new project or new engagement with a country, we often find that there are incredibly rich data sets often hidden away in some ministry or some private company or some UN agency that we haven't spoken to. And it's that often the solutions can be there already and it's that fragmentation and it's bringing people together and in a room together and being driven by the problem that we're trying to fix. So, I think the more that the more that IDAC and UNICEF and other UN agencies can link up and be on the lookout for data sets that are often collected for some other reason but have. Incredibly valuable data for distribution of vaccines or something that's actually collected for an economic survey. The more we can link that up, the better and also make again, my kind of final point of making the pace for the value of data. We're seeing these trends of funds being cut for sensitives, funds being cut for surveys. So, the more we can make that case, we always, unless we can do that, we'll be struggling even more and having these kind of conversations again. Sorry, Sarchil, you had your hand up as well.

Sarchil Qader

[53:17] Yeah, sorry, I just wanted to add that I think there is a huge need among the UN agencies and government and data providers to come together. To work together to understand and make sure that the data that they are producing are interoperable. And because currently we are actually struggling with lack of harmonizations within those data sets that we have. There is, as Andy said, there's a reach of like... huge amount of data out there, but because all of those data has been collected in differently within different formats, different structure. So, it is very difficult to work with them. And also, I think now with the advance in technology and, GPS kind of record and digitisations, I think it will make sense that, the future data collection has always have a sort of like georeferenced kind of component to it that we can georeferenced those data easily into the real world application and that can make it easy for data analysis. I just want to add that to what Andy said. Thank you.

Michael Silberman

Thank you. We're in our final round here. So, Eimear, give us a closing comment if you will. I see your hand is up.

Eimear Cleary

[54:42] That's a lot of pressure. Well, just to add to, yeah, to add to what Andy and Sarchil were saying, I think in as much as possible, I know that IDAC conduct travel surveys and amongst children in particular. So, I mean, adding one or two questions about mobile phone ownership and app use could give us a lot of insight into the digital data that we do currently use if we have that kind of data collected in different contexts.

Michael Silberman

Thank you, well. On behalf of IDAC and the Innovation Working Group, I just want to thank all of our colleagues here from WorldPop for providing this state of play, this bird's eye view that we were hoping to get. You all are kind of at the tip of the spear on much of the research about what's and the application of what is possible and what's not. So, this was. I think we all appreciate the amount of information you were able to share with us in such a short period of time. So, thank you again for doing this. And to the rest of our colleagues from the community, this is the third in the series. We'll be sharing the slides. We'll be sharing the recording. We'll

send an e-mail out when that's ready. It'll be on the data for children on the move.org website, just like the other ones. We will look forward to seeing you at our next session in the series. If you're not receiving, we're forwarded this invite. Please make sure you sign up as well for the e-mail list on that on the data for children on the move.org website so that you get the invitation for our next session in the series with that. Good day. Thank you all. Wishing you good rest of your mornings, afternoons, evenings, wherever in the world you may be today. So, thanks again, WorldPop colleagues. Much appreciated. Thank you. Bye, everyone.