Making the Invisible Visible through Statistics

The High Commissioner just told us to get out of our comfort zones. I have hence decided to focus my talk on Statistics, Science and Sewers. Do I have your attention?

We have collectively agreed “To leave no one behind”. It is an aspirational commitment and the right thing to do. But to deliver, we need to know who we are talking about.

We may have sufficiently strong population statistics and model estimates to confidently say that we are 8 billion people on this planet.

But there are large swaths of this population that are – by and large – invisible. We don’t know what their level of income or poverty is. What their vaccination rates are. What their educational attainments are. Their skill levels and other human capital that they bring. A large part of this invisible population are among the 103 million displaced.

This Dialogue is about finding solutions, so how can we fix this problem?

I will focus on Statistics and science.

Let me start with statistics, and a simplistic overview of how we produce statistics.

First, we have censuses that collect basic information about how many people live here and their demographic profiles. Age, gender, residency, etc.

Second, national statistical offices and government agencies conduct surveys, like the Household Budget Surveys, Demographic and Health Surveys, Labor Force Surveys. They are rigorous, provide rich sets of data but are costly. They help governments guide their investments in schools, hospitals or fire stations.

Together, the census and the national surveys make up the statistical bedrock of our societies, and the data they generate is used by both public and private decision-makers.

What you may not know, however, is how little of this kind of information we have on the forcibly displaced. Too often they are not included in censuses nor in nation-wide surveys. They are statistically invisible, or they live in the statistical shadows of society.

The good news is that we have made considerable progress in the past years to improve the amount, the quality and the accessibility of socioeconomic data on the forcibly displaced.

We now have about 550 datasets from over 80 countries in UNHCR's new microdata library. Three years ago, that number was 0. We have used cutting edge technology – science! – like machine learning and

natural language processing to identify relevant data from the vast sea that exists.

The bad news is that blind spots remain. We are currently piloting a data mapping exercise for the top ten low-income countries that host the largest number of refugees. Early results tell us that while the situation is improving for refugees, real data gaps – blind spots – remain in certain geographical areas, like the Sahel, DRC and the Sudans, and particularly on IDPs.

The way to fill the gaps – the solution – is with statistics and [data] science. To build a solid information infrastructure. And just like the sewage system in any community, it requires upfront investments. We need common standards, tools for data collection, and the capacity to maintain them.

And just like waste disposal, once it works it provides a critical societal function that saves taxpayers lots of money. Sewers and statistics also have in common that as long as it works, few notice. But when it doesn't, it's a mess!

[And just to reassure you, I am not alone to use this metaphor. Perhaps the most famous data journalist alive uses it in his book "The Data Detective".]

In short, we need to invest in our common information infrastructure by supporting inclusive censuses and national surveys and make use of what data science offers.

To illustrate with a concrete example, let me tell you about Uganda's bureau of statistics, UBOS. As they were planning the next Demographic and Health Survey, the government wanted to include the 1.5 million refugees who live in the country, and make good on a

statistical inclusion pledge that they made at the Global Refugee Forum in 2019.

Considering the technical challenges and the costs, they turned to the World Bank, the Joint Data Center, and UNHCR for support. The survey is currently underway.

The data from this exercise will enable the government to better plan its public health services, which they already generously extend to refugees. To statistically include a population of 1.5 million people will address a statistical blind spot and significantly improve the allocation of resources.

Tomorrow, we will be launching an innovation lab on statistical inclusion, asking more countries to do that same. We are inviting host

countries and development partners to make a public commitment to invest in statistical systems and to ensure that the forcibly displaced are made visible, and thereby contributing to our overall ambition of leaving no one behind.

To achieve this, we need to work together. The Joint Data Center stands ready to do its part.

Thank you