Leaving no one behind: Appropriate aggregation and right disaggregation

Who is left behind?

How does it feel to be part of a particular subgroup of a given population? The SDGs ambition of leaving no one behind requires timely and high quality disaggregated statistics that provide answer to this fundamental question about citizens no matter who and where they are. The 2030 agenda for sustainable development should achieve the "leave no one behind" goal by empowering the furthest left behind population groups, giving them voice and widening their choices. However, the first step, prior to any action, is to identify and acknowledge the most vulnerable, discriminated against and excluded groups of people and understand their lives. In other words, in the context of the need for disaggregated statistics for implementation of the 2030 development agenda, the starting point is identifying "people" who are likely to be left behind rather than "numbers" that need to be disaggregated. After all, it is government policies and programmes prioritizing and targeting those population subgroups for interventions that are likely to improve the lives of the most deprived part of the society, and disaggregated statistics can inform the formulation of such interventions.

There are five different criteria* that may help us better identify our target population subgroups that are likely to be left behind:

1 Hard to reach

Sub-groups of population that are difficult to target for a variety of reasons such as being small (in the minority) or having specific characteristics such as illness, occupation, etc.

2 Hidden population

When public acknowledgement of the population is potentially threatening for the members of the sub-group. Size of these population groups is often unknown and strict privacy issues are a concern in identifying them.

3 Excluded, marginalized,discriminated

Though the three groups are different, all share the same characteristic, they are often "known" but "ignored" in one way or another. Examples: certain ethnic groups, certain age groups, sex, occupation, religious minority groups.

4 Vulnerable sub-population groups

A sub-group that is potentially in a disadvantaged position due to its socio-economic situation. Examples: uninsured, low income, slum, or elderly groups.

5 Geographically disadvantaged

Sub-populations that live in an unfortunate situation due to geographical conditions such as harsh climate, remote and hard to access locations, poor infrastructure.

** The groups are not mutually exclusive and one person may be identified in more than one group*

Every individual who belongs to one or more of the above subgroups may be at the risk of being left behind and issues affecting his/her life have to be addressed in policy interventions. Data collection methods and tools may be designed to produce statistics at "appropriate level of aggregation" that inform policy programmes. For instance statistics may be produced within groups identified by sex, specific age groups and labour force characteristics to target female youth who are unemployed. However, it is often not feasible to target all subgroups in the data collection phase. For instance HIV positive population from an ethnic minority group residing in a remote area may belong to the first three subgroups (hidden population, marginalized, hard to reach) and difficult to capture by conventional data collection methods and tools. In this situation, there is a need for disaggregating statistics after they are produced for broader objectives to generate information about a specific target subgroup. Identification of needs and issues of the target subgroups prior to disaggregation is vital for coming up with "right disaggregated statistics".

The five criteria introduced above provide us with a framework to identify interlinkages between characteristics that together identify a certain group of people that are targeted by policy interventions and has to be basis for disaggregation strategy.

Ideally, identification of the sub-groups that need to be prioritized in national policies, and acknowledgment of existence of the excluded population groups should help us to produce statistics at an appropriate level of aggregation that ensure no one is left behind. Data collection has to be informed by and designed to address the needs for informing target interventions. For instance, appropriate data collection procedures and tools should be applied to ensure representation of members of target groups and their issues. The same is required for data analysis.

Traditionally, most of the official statistics are produced from data collected through sample surveys that are designed to mainly produce aggregates for the major administrative divisions within countries. Geographical disaggregation, though important, can only show one dimension of deprivation or disparity. Addressing the "leave-no-one-behind" focus of the 2030 Agenda requires aggregation/disaggregation by other characteristics of population groups, including lower levels of administrative divisions.

Data disaggregation in support of the implementation of the SDGs

The proposed SDGs indicator framework provides two types of guidelines for disaggregation: (a) when disaggregation has to be standardized at the global level and requires international comparability (such as specific age groups, types of disease, etc), it has been embedded into the indicator structure; and (b) when disaggregation has to be done based on population characteristics that may vary across countries depending on their policy priorities and other circumstances, the framework provides a minimum list of characteristics for which disaggregation is most desirable (sex, age, income, race, migratory status, disability, and geographic location). The framework also encourages countries to identify the level of disaggregation beyond what is proposed in global SDGs indicator framework as required for monitoring their national policy programmes. The Inter-agency Expert Group on SDG Indicators (IAEG-SDGs) is exploring options for developing a consolidated tool to guide national statistical systems to put in place the building blocks for identifying the most vulnerable populations and for producing required disaggregated statistics.

In the Asia-Pacific region, expert dialogues** have identified key steps required for national statistical systems in developing disaggregation strategy for production and dissemination of official statistics:

- Governments to review their national legal and policy frameworks in light of the international development priorities (including SDG framework) to identify target population groups (likely to be left behind) and issues to develop a standard disaggregation strategy most desirable for their country.
- Advancing technical capacity and access to methodologies by national statistical systems to maximize use of existing data sources, including administrative data, for producing disaggregated statistics and increasing access to and use of micro-data.
- Supporting national efforts to improve the production and dissemination of analytical work focused on disparity analysis.
- Increasing effective user-producer dialogue to make sure that statistical system is producing relevant and essential evidence for monitoring socially inclusive development.

The process of developing a disaggregation strategy is an iterative and interactive process that should be mainstreamed into the statistical production system. It starts by active communication between users and producers to identify sub-groups that are to be prioritized by policies and programmes. At this stage, it is important that a common language is developed through regular communications; a language that talks about "people" rather than "numbers". Through this process, target population groups are identified, and then producers of data and statistics have to re-think design, tools, collection methods and procedures, and formulate dissemination strategy that produces appropriately aggregated statistics.

***http://www.unescap.org/resources/report-workshop-sex-disaggregated-data-sdg-indicators-asia-and-pacific-what-and-how*

Opportunities for production of relevant disaggregated statistics

The good news is that modernization of statistical business processes and diversification of sources of data used for compilation of official statistics bring new opportunities for producing more relevant disaggregated statistics. Increased access to micro-data, integration and linkages of different data sources, increasing use of Big data in producing official statistics, and enhanced tools and capacity for applying statistical methods for disaggregating statistics are major steps that national and international statistical communities are taking in providing data support to the "leave no one behind" vision of the SDGs.

Increased access to micro-data

In-depth understanding of who is left behind and what interventions are effective often requires detailed analysis of micro-data. Access to micro-data by users, including academia and civil society, can improve availability of statistics about furthest left behind groups of people by generating right aggregation and when necessary appropriate disaggregation.

Data integration

Statistical integration has to happen at all three levels of source, production and dissemination. However, integrating data from different sources including mainly from surveys and censuses, administrative registers and new sources of data is key to maximize data disaggregation and in depth analysis of issues that affect lives of people in various population groups.

Harnessing the power of Big Data

Use of unstructured data that are not necessarily produced for statistical purposes is a challenging task, but at the same time an unprecedented opportunity for producers of official statistics to generate information on aspects of life that are not captured by conventional data collection procedures. Big data such as data produced by social media, mobile phones, scanners and image analysis, if harnessed efficiently and within quality assurance framework, can provide a rich source of data about population groups that are likely to be excluded from traditional data sources.

Using advanced statistical methods

Capacity of national statistical systems to apply statistical techniques such as small area estimation for disaggregating statistics by a combination of desirable population characteristics has enhanced over the past years. Increased availability and access to auxiliary information, geospatial data and micro-data, in particular from administrative sources and statistical registers, are major factors that facilitate application of more sophisticated statistical methods for producing disaggregated statistics.