

Ensuring no one is left behind: A human rights-based approach to data

Why are averages not enough?

Official statistics have traditionally focussed on national averages, whether in the field of economics or to aid policymaking on the provision of healthcare, education or social services. However, averages by their very nature mask disparities and are inadequate as a sole measure of progress. If our aim is genuinely to leave no one behind, we must learn the lessons of the MDGs and ensure that the indicators chosen to measure progress towards the SDGs include a clear focus on the most marginalized and vulnerable members of society. This means that SDG data collection should capture not only national averages or aggregate statistics, but also the situation of the most disadvantaged or deprived, as well as the inequalities among social groups.

What groups do we need data for?

Current censuses and household surveys all too often exclude the most vulnerable or marginalised. This may be due to design limitations, for example when surveys conducted at a person's home exclude the homeless or when the person conducting the survey cannot communicate with a member of a linguistic minority. But it can also be due to fear and stigmatisation of identifying as a member of a particular minority, or mistrust as to the purposes for which personal data may be accessed or used. A true data revolution must therefore include innovative approaches to reaching the marginalised groups who are currently invisible in official statistics.

Relevant data should be disaggregated by all grounds of discrimination prohibited by international human rights law, as enshrined in the 1948 Universal Declaration of Human Rights and subsequently elaborated upon by international human rights mechanisms. Such grounds include ethnicity, sex, age, income, geographic location, disability, religion, migratory or displacement status, civil status, sexual orientation and gender identity. While some disaggregation should be common to all countries and follow standard definitions, such as that on sex, age or disability, the precise categories to be included under grounds such as ethnicity, geographic location and religion will vary according to national circumstances. However, differences in the make-up of populations cannot be used as a justification not to measure the progress of the most marginalised towards the SDGs.

The categories by which data should be disaggregated should be determined in an inclusive, participatory process at the national and sub-national levels, with the direct involvement of minority groups themselves. The Committee on the Elimination of Racial Discrimination (CERD) has made clear that identification as a member of a particular ethnic group "shall, if no justification exists to the contrary, be based upon self-identification by the individual concerned." This principle should apply across the SDGs. Groups themselves should determine how they are identified on any list of options, and it should be possible for individuals to choose not to identify as a member of any group, or to enter their own choice of identity. In some cases, proxy indicators such as language spoken at home may be more appropriate than direct questions about group identity if these are likely to reinforce divisions, but such choices must be made with the full participation of groups themselves.

Is disaggregation always the best way to detect inequalities?

No, sometimes other types of data collection will be more appropriate. Inequalities may be detected, for example, through disaggregation of data, by calculating the share of public expenditure spent on public services for each group, or through targeted surveys.

Where the size of a group is very small compared to the size of the whole population, too few members of the group would be captured by a general survey to allow for general conclusions about the group to be drawn. This may be overcome by weighting the survey with the aim to increase representation of specific groups, or by conducting a separate survey for the targeted group, which can then be compared to the national survey.

Where a group is victim of on-going or historical, direct or indirect discrimination by authorities or even other members of society, members of the group may not be willing to identify as such. This risk may be reduced by ensuring clear and transparent safeguards on data storage to protect the privacy of data subjects. However, even this may not be enough, and collection of data through civil society organisations or service providers that enjoy the trust of the most marginalised groups may produce more reliable results. This will require new partnerships between National Statistics Offices (NSOs) and civil society organisations (CSOs), and capacity strengthening within both CSOs and NSOs.

What is the role of big data in ensuring no one is left behind?

In addition to more ‘traditional’ data sources such as socioeconomic statistics, administrative records, household surveys and expert opinion, new potential sources include social media, online or mobile-based crowdsourcing, automated content analysis of a large quantity of online media, satellite data, and even data on spikes in mobile telephone usage. Such data do not in general suffer from the time-lag of traditional data sources, where there is often a long gap between collection and publication, and so can provide vital information in crisis situations. However, their usefulness in on-going monitoring is unclear, and it is telling that none of the currently proposed indicators for the SDGs involve “big data.”

While big data can be useful, it is vital to ensure representation of all members of society in measuring progress towards the SDGs. Many new data sources rely on literacy and access to an internet connection or mobile telephone, and so may exclude the most marginalised groups. In collecting such data, it is generally impossible to determine or ensure coverage of the total population. Such sources should therefore be used with caution, and only in tandem with other data sources that have adequate safeguards in place to ensure representation of all members of society.

What kind of data revolution do we need?

Member States have indicated on numerous occasions, including in the Open Working Group deliberations and final report, that no target should be considered achieved unless met for all population sub-groups so that no one is left behind. There is a need for a ‘data revolution’ for data disaggregation and targeted data collection to capture the situation of the most disadvantaged groups and the groups affected by discrimination, including multiple and intersecting forms of discrimination. This revolution must involve not only experts from the statistical community but also civil society organisations, national human rights institutions, service providers and, most importantly, marginalised populations themselves.