Statistical Commission   Background document
Fifty-second session   Available in English only
1–3 and 5 March 2021
Item 3(a) of the provisional agenda
**Items for discussion and decision: Data and indicators for the 2030 Agenda for
Sustainable Development**

Practical Guidebook on Data Disaggregation for the SDGs

Prepared by the Asian Development Bank (ADB)
and the United Nations Statistics Division (UNSD) with input from
the Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs)

Practical Guidebook on Data Disaggregation for the SDGs

Current draft as of January 2021

Foreword

To build a road map for people and the planet and to ensure sustainable social and economic progress worldwide, the 2030 Agenda for Sustainable Development was launched in 2015 as a universal call to action for ending poverty, protecting the planet, and ensuring that all people enjoy peace and prosperity. Espousing the "leave no one behind" principle, the United Nations Member States commit to eradicate poverty in all its forms, end discrimination and exclusion, and reduce the inequalities and vulnerabilities. To understand which population groups are left behind and design effective policies, it requires that the indicators monitoring the Sustainable Development Goals (SDG) should be disaggregated by income class, gender, ethnicity, geographic location, disability status, migration status, and other relevant dimensions. However, disaggregation of SDG indicators imposes significant data requirements and operational challenges for national statistical systems.

In many countries, development data are compiled as national, provincial, or city averages. Although such type of aggregated data painted 'big pictures' revealing where the whole society stands, in general, with respect to specific development targets, it did not provide adequate information on which segments of the country's population made significant progress or lagged in terms of development. From a policy perspective, the lack of granular data is problematic because there are limited data to guide the design of intervention programs meant to appropriately target vulnerable segments of society. On the other hand, where disaggregated data are available, evidence suggests that the most vulnerable groups were more likely to have benefitted disproportionately less from the development process than the rest of the population, contributing to widening inequalities within countries. Worryingly, during periods of uncertainty such as the ongoing COVID-19 pandemic, sparse data suggest that the poor and vulnerable groups also tend to be at greater risk of social and economic exclusion.

In 2017, the Asian Development Bank (ADB) designed a technical assistance project called Data for Development, which aims to strengthen the capacity of national statistics offices (NSOs) in the Asia and Pacific region to meet the increasing data demands for SDG monitoring and effective policymaking. One component of the project focuses on providing guidance on how national statistical systems can enhance the compilation of disaggregated data for development. In relation to this initiative, statisticians from ADB's Statistics and Data Innovation Unit within the Economic Research and Regional Cooperation Department worked with the Statistics Division of United Nations Department of Economic and Social Affairs (UN DESA) and other development partners to draft a practical guidebook that can be used by staff of national statistical offices and other organizations that compile SDG and other data for development. In particular, the guidebook provides tools to collect, compile, analyze and disseminate disaggregated data. It provides background materials regarding issues and experiences of countries regarding data disaggregation for the SDGs. This guidebook is intended for statisticians, as well as other analysts from planning and sectoral ministries involved in the production, analysis, and communication of disaggregated data.

The publication team was led by Arturo Martinez, Jr, under the overall direction of Elaine Tan. The preparation of this guidebook began as a series of notes from the International Workshop on Data Disaggregation for the SDGs, organized by the Statistics Division of UN DESA, in Bangkok in January 2019, which were initially summarized by ADB consultant Jose Ramon Albert with the assistance of Arturo Martinez Jr., and subsequently served as one of the main references in finalizing the guidebook. ADB consultant Margarita Guerrero wrote the guidebook, with the

guidance and inputs provided by UNSD's Heather Page, Yongyi Min, and ADB's Arturo Martinez Jr., Mildred Addawe, Marymell Martillan, Joseph Bulan, and Ron Lester Durante. United Nations agencies provided input on tools and resources, which included the World Health Organization, UN Women, UN ESCAP, and UNICEF. The Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) also provided comments on the drafts and encouraged widespread dissemination of the Guidebook, including as a background document to the 52nd session of the Statistical Commission in March 2021. Arman Bidarbakht Nia (UNESCAP), Sara Duerto Valero (UN Women), Ahmad Reza Hosseinpoor (WHO), and Francois Fonteneau (PARIS21) all contributed to works that were used as inputs and references for this guidebook. Rose Anne Dumayas provided operational support through the course of the project. The cover of this guidebook was designed by Rhommell Rico. Person XX provided editing services, ensuring coherence and consistency. Manuscript editing was performed by Person XX, while the publication's layout, page design, and typesetting were carried out by Person XX.

The goal of this practical guidebook is to share knowledge that can improve the capacity to produce, analyze, and communicate disaggregated data for the SDGs. It is designed to be an initial publication that features analytical tools on producing and using disaggregated data, based on experiences of countries, the IAEG-SDGs, and international and regional organizations. Succeeding updates of this guidebook will be subsequently available on the following website (https://unstats.un.org/sdgs/).

The authors hope that this document represents a useful reference for delivering high-quality, granular, and cost-effective data for SDG monitoring.

Introduction

Background and Rationale

"Leaving no one behind" (LNOB) is a central cross-cutting focus of the 2030 Agenda for Sustainable Development to which United Nations Member States committed to—

> *"As we embark on this great collective journey, we pledge that **no one will be left behind**. Recognizing that the dignity of the human person is fundamental, we wish to see the Goals and targets met for all nations and peoples and for all segments of society. And we will endeavor to reach the furthest behind first."*[1]

It calls for all targets to be reached by everyone and **asks for detailed disaggregation of measures of progress by groups** to ensure that progress towards these targets is reached by all. In addition, some goals are also directly concerned with those *currently* left behind; that is, the ***vulnerable populations***. For example, the goals to eliminate poverty, hunger, and preventable child mortality. The goal on (income) inequality calls for a reduction in disparities, where improvements in the well-being of those left behind will be an important strategy. And, the strengthened gender equality goal focuses on women and girls who are left behind[2].

Ensuring that these commitments are translated into well-reasoned evidence-based policies and corresponding effective actions requires a precise understanding, definition, and identification of the target populations. To properly measure progress in attaining the targets and goals, data needs to be collected and statistics need to be generated for clearly defined target populations.

To this end, the United Nations Statistical Commission (UNSC)[3], charged with developing the overall measurement framework and indicators for progress monitoring of Agenda 2030, embraced an overarching principle of data disaggregation in the development of the Global Indicator Framework for the Sustainable Development Goals (SDGs) and target--

> *"Sustainable Development Goal indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics, in accordance with the Fundamental Principles of Official Statistics."*[4]

Disaggregation of SDG indicators along these dimensions imposes significant data requirements and operational challenges for national statistical systems (NSSs). Thus, the UNSC created the Inter-Agency and Expert Group on Sustainable Development Goal Indicators (IAEG-SDGs)[5] to develop and implement the global indicator framework for the Goals and targets of the 2030

---

[1] Para 4. United Nations. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. https://sustainabledevelopment.un.org/post2015/transformingourworld
[2] For an overview of issues on understanding the concept of LNOB and problems involved in its operationalization, refer to: Stephan Klasen and Marc Fleurbaey. 2018. Leaving no one behind: Some conceptual and empirical issues. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/CDP_BP44_June_2018.pdf
[3] The UNSC is the highest decision making body for international statistical activities, responsible for setting of statistical standards and the development of concepts and methods, including their implementation at the national and international level. https://unstats.un.org/unsd/statcom/
[4] United Nations General Assembly Resolution 71/313. https://undocs.org/A/RES/71/313
[5] Information on the mandate, membership and work programme of the IAEG-SDG. https://unstats.un.org/sdgs/iaeg-sdgs/

Agenda. Implementing the indicator framework includes providing necessary statistical standards and tools to assist national statistical systems in the production of indicators with the recommended levels of disaggregation. This work-in-progress, introduced in Chapter 1 of this Guidebook, has so far resulted in identification of *dimensions* or characteristics by which indicators are to be disaggregated (e.g., sex, age, disability) and corresponding *categories* (e.g., male/female for the sex dimension). The IAEG-SDGs has also defined a *minimum disaggregation set,* a set that includes all disaggregation dimensions explicitly referenced in the target or indicator name, prepared an overview of standards for data disaggregation and identified *policy priorities* of different vulnerable population groups to advice on future focus of data disaggregation[6].

This Guidebook is designed as a knowledge resource that brings together existing statistical standards and tools that have been and can be utilized by countries in taking steps to provide disaggregated data for compiling SDG indicators as defined by the IAEG-SDG work on data disaggregation. The Guidebook also demonstrates analyses highlighting disparities and inequities in key policy areas made possible by availability of disaggregated data. The intention is to provide a resource that can be referred to as a starting point for undertaking the statistical work required.

Purpose of the Guidebook

For whom

The Guidebook is primarily intended for statisticians and data analysts of national statistical offices and planning, and sectoral ministries involved in the production, analysis, and communication of data and statistics in support of inclusive sustainable development, particularly the monitoring of progress in achieving the SDGs. Researchers, academia, civil society organizations, private sector and information providers and users in the national data ecosystems may also find the Guidebook useful and relevant.

For what

This Guidebook is intended to provide information on existing statistical sources, methods, tools and current initiatives that address some of the key issues that need to be considered in the production and analysis of data needed for generating disaggregated statistics and indicators and the reporting and communication of such, in order to be able to:

- Better understand the concept of disaggregation as applied to data, statistics and indicators and the role of disaggregation in formulating, monitoring and achieving national development goals, including the SDGs

- Better understand and respond to the policy-data nexus critical to inclusive and "leave-no-one-behind (LNOB)" development as espoused by Agenda 2030

---

[6] The compilation of data disaggregation dimensions and categories and the policy priorities are living documents and will be updated as new information is received. More detailed information are available here: https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/

- Produce data needed for estimating SDGs indicators with the specified disaggregation dimensions for indicators for which methods and tools have been developed and demonstrated

- Use data produced to generate disaggregated statistics and SDG indicators and carry-out multi-dimensional data analyses that support policy formulation, analyses and monitoring of the achievement of SDGs

- Effectively present and communicate disaggregated statistics and SDG indicators and the results of multi-dimensional analyses to target audiences

- Strengthen institutional and national statistical system capacity for production, analyses and use of disaggregated SDG indicators in areas where most needed

How

In this Guidebook, the term "**disaggregated data**" refers to data that can be used to generate statistics and indicators for population groups defined by (or disaggregated by or broken down further into) one or more dimensions or characteristics; commonly-- sex, geographic areas, age.[7] The results are referred to as disaggregated statistics or indicators. The entire process is referred to as *data disaggregation*.

In the context of the LNOB principle of the 2030 Agenda, the need for disaggregated data is to be able to identify vulnerable groups or populations that are most likely to be left behind, understanding the factors that keep them in or move them out of that position and to report on the indicators for these groups in order to monitor their progress in achieving the development targets and goals.

The Guidebook provides information and guidance on applying sources, methods and tools for data disaggregation and the analysis, use, dissemination and reporting of the resulting disaggregated statistics and indicators. The information is gathered from various sources, including-- methodological briefs, guidance notes and, where available, internationally recommended guidelines; work of the various task teams of the IAEG-SDGs; publications of the United Nations (UN), Asian Development Bank (ADB), and other development partners; illustrative examples from country work gathered from presentations at regional and international workshops featuring disaggregation for SDGs indicators; tools developed and utilized by international development organizations.

The main topics covered are:

- Concepts and definitions relating to the data disaggregation process and their mapping to SDGs-related dimensions and priorities, based on the IAEG-SDG work on data disaggregation
- Integrating policy demands on inclusive and LNOB development with data and illustrative applications of related tools

---

[7] A related terminology used in the literature is "granular" data which represents the idea of data about smaller chunks or pieces of a larger population.

- Sources of disaggregated data: description, illustrative uses and summary of strengths, potential and limitations
- LNOB approaches to data analysis: policy-data nexus in leaving-no-one-behind and illustrative applications of multi-dimensional analysis and related tools
- Disseminating and communicating disaggregated statistics and indicators: facilitating access to and utilization of SDG-related data; presentation and communication approaches and tools for enhancing understanding and use of disaggregated data in monitoring progress in achieving progress in LNOB in the SDGs
- Knowledge resources for addressing capacity development needs in producing disaggregated data for generating SDGs indicators

Structure of the Guidebook

The Guidebook covers the following topics:

Chapter 1- Discusses the concept of and SDG indicator requirements for "leaving no one behind" and the connection to the statistical concept of data disaggregation. The chapter introduces the disaggregation dimensions and categories and related existing statistical standards.

Chapter 2- Discusses the importance of defining the relevant and appropriate target populations as clearly as possible in terms of the data dimensions explicitly or implicitly stated in development goals and targets and presents tools that can facilitate the identification of relevant development policies and matching data needs and policy needs. The chapter illustrates the use of disaggregated data in developing nuanced policy to address different vulnerable groups a

Chapter 3- Explains and illustrates known strengths, potentials and limitations of statistical sources of disaggregated data considered in the Guidebook, namely: censuses; household sample surveys; administrative registers; data generated in the application of small-area estimation models; big data, geo-spatial data and related data sources; and data produced by applying methods for integrating data sources.

Chapter 4- Explains and illustrates basic methods of generating SDG indicators disaggregated by the specified dimensions and carrying out equity and disparity analyses and some tools that have been developed to do so. The chapter provides illustrative examples of SDG indicators disaggregated according to recommended dimensions as produced by countries and international organizations.

Chapter 5- Illustrates practices in reporting, communicating and enhancing use of disaggregated statistics and analyses, including reporting on SDGs indicators, data visualization and multi-level analyses.

Chapter 6- Identifies key issues on why required disaggregated data are often not available and provides information on knowledge products and ongoing capacity development initiatives that may help national statistical systems to address these issues.

Source references

A list of knowledge resources that provide more details on the topics covered is provided at the end of each chapter. The illustrations and practices are primarily drawn from various regional and global workshops organized by United Nations agencies covering topics on data disaggregation for the SDGs held during the years 2018 – 2020, including:

- International Workshop on Data Disaggregation for SDGs, 28-30 January 2019, Bangkok (co-organized by Asian Development Bank)

- Second meeting of the Statistical Coordination Group for the 2030 Agenda in Latin America and the Caribbean: Disaggregated data for regional monitoring of the SDGs, 3-5 September 2019, Quito

- ESCWA Regional Workshop on Data Disaggregation for SDGs Indicators, 19-21 November 2019, Istanbul

- Counted and visible: Global Conference on the measurement of gender and intersecting inequalities, 26 February 2020, New York

Future updates

The Guidebook is designed to be an initial publication in a series of tools on producing and using disaggregating data, based on the work of various countries, IAEG-SDGs, and international and regional organizations. As more updated tools and references become available over time, the Guidebook will be updated too. The updated Guidebook will be available on the following website (https://unstats.un.org/sdgs/).

Chapter 1. Data Disaggregation and the SDGs Indicators

Overview

It is highlighted in the 2030 agenda that no one should be left behind. For this to happen, data to measure progress in meeting the targets must provide a level of granularity or *disaggregation* that makes visible the most vulnerable and marginalized people. Depending on their sex, age, education, income, location, and other intertwining factors, people face different constraints. Disaggregated data allows more effective and efficient policies and interventions. This information can show what works well, what needs more attention, and who are being served by policies, programs, and projects.

The Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) has compiled existing standards, thoughts, and ideas on data disaggregation for the minimum set of data disaggregation dimensions. The IAEG-SDGs classified the disaggregation dimensions of monitoring the SDG indicators into three categories – minimum set, other dimensions in addition to the minimum set, and dimensions for which international custodian agencies may introduce in the future in addition to the first two categories.

The task of producing the levels of disaggregation of SDG monitoring indicators needed to better understand the situation of the most vulnerable and marginalized people is a difficult one. National statistical systems need to prepare a strategy to improve the quality, quantity, and availability of disaggregated data and statistics. That strategy should include financing to support capacity to produce and use these.

1.1   What is data disaggregation and why is it important?

In this Guidebook, the term *data disaggregation* is used to mean:

> *"… the breakdown of observations within a common branch of a hierarchy to a more detailed level to that at which detailed observations are taken. With standard hierarchical classifications statistics for related categories can be split (disaggregated) when finer details are required and made possible by the codes given to the primary observations."*[8] [9]

In contrast, *data aggregation* is

> *"… the combination of related categories, usually within a common branch of a hierarchy, to provide information at a broader level ... With standard hierarchical classifications, statistics for related categories can be grouped or collated (aggregated) to provide a broader picture …"*[10]

'Big pictures' do not always portray the 'full picture'

---

[8] UN Glossary of Classification Terms. https://unstats.un.org/unsd/classifications/bestpractices/glossary_short.pdf
[9] The term "granular" is often used interchangeably with "disaggregated", with finer granularity meaning more detailed levels of disaggregation.
[10] UN Glossary of Classification Terms. https://unstats.un.org/unsd/classifications/bestpractices/glossary_short.pdf.

Compelling human interest stories often spur collective action. However, simply telling stories without data limits assistance to the few subjects of these stories. Data, particularly official statistics, can be used as leverage to speak on behalf of the ones left behind. These statistics should be able to identify who the vulnerable, disadvantaged, marginalized, or socially excluded are, how many they are and where they are located. These are the minimum information needed for policymaking and designing effective intervention programs for marginalized groups. These information enables societies to hold governments accountable when looking at why segments of the population get left behind in the development process.

However, official statistics have been generally made available in the form of aggregated data. When data are aggregated, they provide a 'big picture'. The focus on leaving-no-one-behind (LNOB) in the 2030 Agenda has highlighted that the 'big picture' does not always portray the full picture. That is, actual living conditions of segments of society such as persons with disabilities, people living with precarious health conditions (e.g., HIV/AIDS), indigenous peoples, undocumented migrants, religious minorities, refugees, the uninsured, the elderly, those internally displaced, or those who are in vulnerable working conditions, among others, are hidden—at times inadvertently but, unfortunately in other situations such information are deliberately suppressed. As illustrated in Box 1-1, being absent in the data masks the extent of deprivation and disparities they face, and further exacerbates their vulnerabilities.

Box 1.1: Disaggregated data provide a more nuanced picture of development landscape

To illustrate how aggregated data can hide important details about the development process, consider two countries, A and B. In country A, the more populous of the two countries, 80 per cent of primary students have achieved a minimum proficiency in reading and mathematics, while in country B, only 60 per cent have attained at least minimum proficiency (Figure 1.1-a). The picture of quality of education in the two countries changes if further information is provided that in country A, half of all children of primary-school age do not go to school and they all come from poor families, while in country B, all children in the same age group are attending school. This would therefore indicate that only 40 per cent of the primary-school age population in country A has minimum proficiency in reading and math (Figure 1.1-b).

Suppose further that 80 per cent of those with minimum proficiency in country A are boys, while in country B, both boys and girls have the same opportunity (Figure 1.1-c). Imagine further if one of these countries faced large disparities between different geographic regions of the country, so that the national figures show an average between some relatively extreme strong and poor cases across the country. In this case, aggregated data conceal the real state of education inequities.
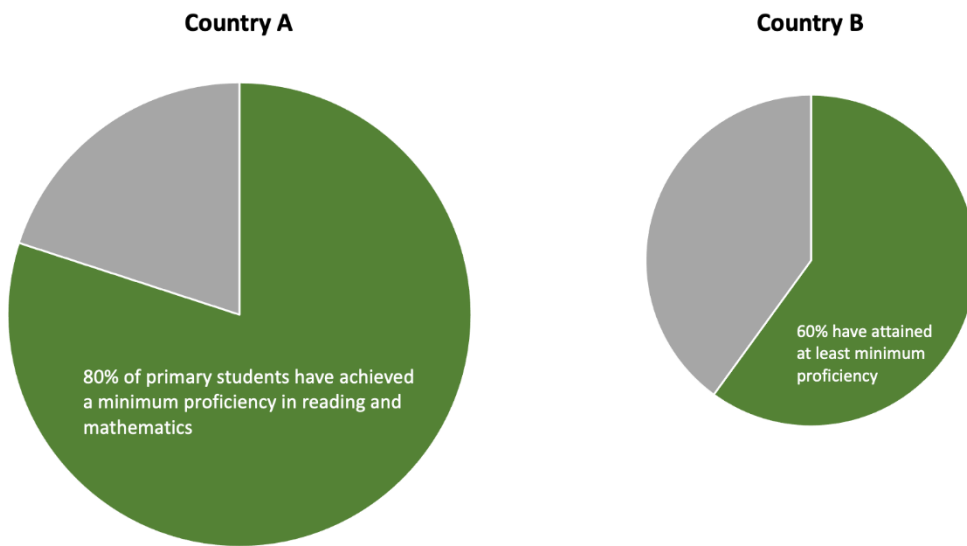

Country A
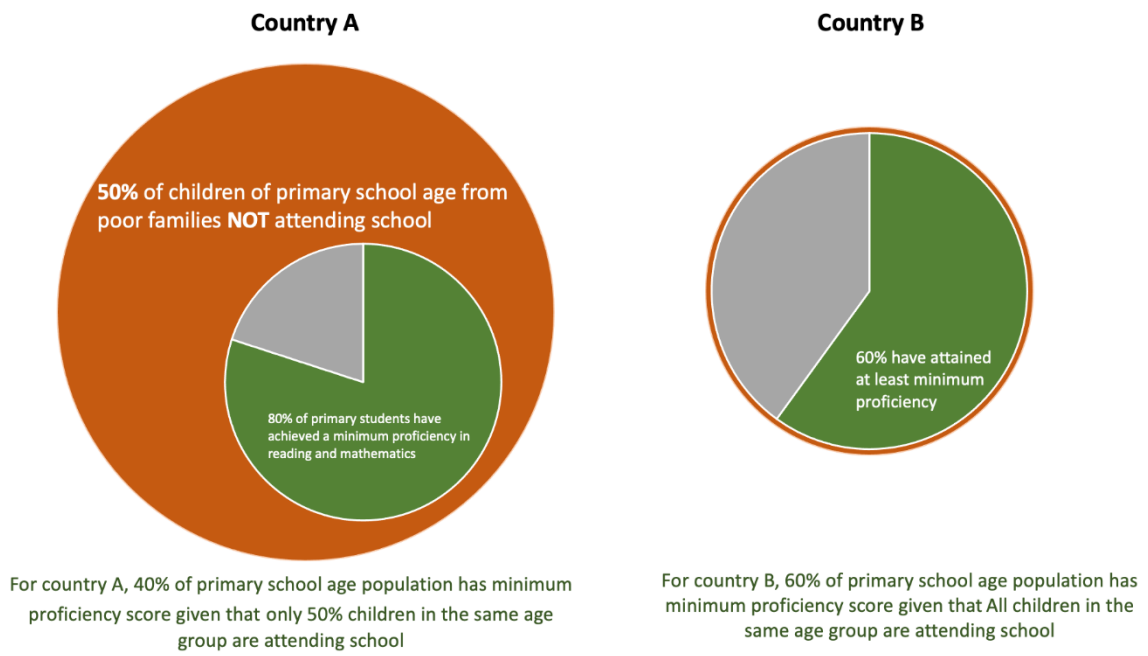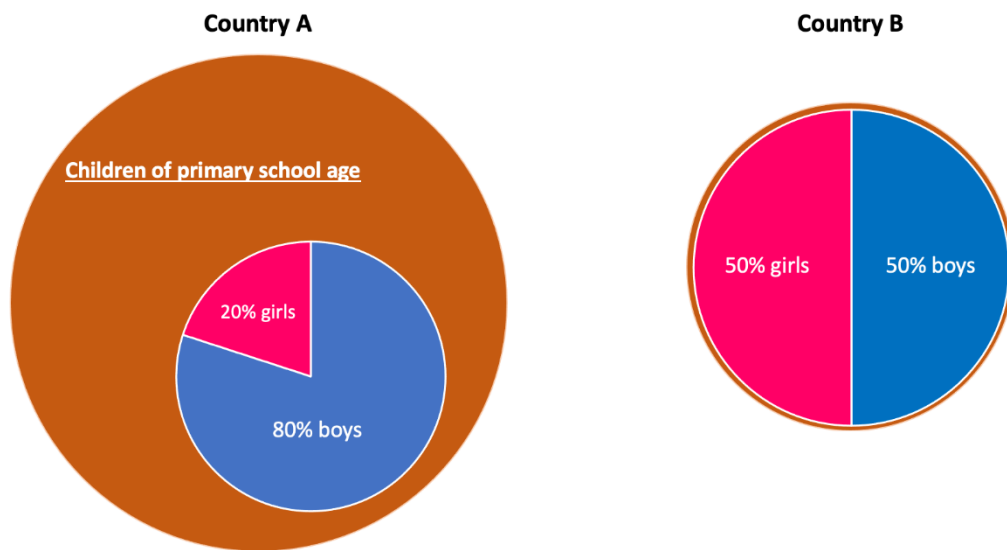

80% of primary students have achieved a minimum proficiency in reading and mathematics

Country B

60% have attained at least minimum proficiency

Figure 1-1-a

Country A

50% of children of primary school age from poor families **NOT** attending school

80% of primary students have achieved a minimum proficiency in reading and mathematics

For country A, 40% of primary school age population has minimum proficiency score given that only 50% children in the same age group are attending school

Country B

60% have attained at least minimum proficiency

For country B, 60% of primary school age population has minimum proficiency score given that All children in the same age group are attending school

Figure 1-1-b

Country A

Children of primary school age

20% girls

80% boys

For country A, 40% of primary school age population has minimum proficiency score given that only 50% children in the same age group are attending school

Country B

50% girls

50% boys

For country B, 60% of primary school age population has minimum proficiency score given that All children in the same age group are attending school

Figure 1-1-c

The same applies to indicators on poverty, decent work, health and sanitation, and women empowerment, etc. when aggregated data provide only big pictures of socio-economic conditions, but do not necessarily portray the full picture, especially with regard to inequities.

Aggregation loses information as aggregates often hide the disparities that exist between population groups. In contrast, disaggregation involves breaking down data into smaller information units. When data are sufficiently disaggregated, multidimensionality and intersectionality[11] of inequalities are better surfaced out and analysed.

Discrimination, geography, governance, socio-economic status, as well as shocks and fragility are the five key factors[12] that can hide vulnerable groups across societies within aggregated data. These factors can intersect, compounding the deprivations, and reinforcing the limitations of people who live in the margins. Consequently, the furthest behind are most likely enduring challenges from multiple and intersecting forms of disadvantages. Even at high levels of disaggregation (world subregions; urban-rural; broad age groups), disparities become visible—as shown in the statistics cited in Box 1-2.

Policies should be geared to respond to various concerns across demographic groups. People face different constraints depending on their sex, age, education, income, location, and other factors. Information derived from disaggregated data enables more effective and efficient policies

---

[11] *Intersectionality* is the interconnected nature of social categorizations such as race, class, and gender, regarded as creating overlapping and interdependent systems of discrimination or disadvantage (Source: Oxford Dictionary)

[12] UNDP. 2018. What does it mean to leave no one behind? A UNDP discussion paper and framework for implementation. https://www.undp.org/content/undp/en/home/librarypage/poverty-reduction/what-does-it-mean-to-leave-no-one-behind-.html

and interventions. Data can show what works well, what needs more attention, and who are being served by policies, programs, and projects.

Box 1.2: Even at coarse levels of data disaggregation, disparities can be revealed

At the start of the SDG implementation period in 2015, 85 per cent of urban residents globally used safely managed drinking water services, compared to only 51 per cent of their rural counterparts. Nine out of ten (89 per cent) births from the richest quintile were attended by skilled health personnel, compared to less than half (43 per cent) among the poorest quintile.

Between 2014 to 2016, the prevalence of undernourishment was significantly higher in least developed countries (27 per cent), landlocked developing countries (23 per cent), and small island developing States (18 per cent) than in the developing regions as a whole (13 per cent). Children from the poorest households are more than twice as likely to be stunted compared to their richest counterparts based on a survey of 87 countries from 2005 to 2014.

About three out of five victims of human trafficking in sub-Saharan Africa (65 per cent) and South-East Asia (57 per cent) are children.

Data from 63 countries surveyed between 2008 to 2012 showed that children from the poorest households are nearly four times more likely to be out of school than children from the richest households.

Youth aged 15 to 24 years old are nearly three times more likely to be unemployed than adults 25 years old and over. Further, youth are also more likely than other age groups to be poor: half of people living in extremely poor households (385 million) were under 18 years.

Across four Asian countries-- Bangladesh, the People's Republic of China, Mongolia and Pakistan-- around 90 percent of women perform unpaid care and domestic work, compared to 31 to 75 per cent of men.

1.2   Disaggregation Dimensions for the SDG Indicators

The initial work on disaggregation by the IAEG-SDGs identifies disaggregation *dimensions* and their corresponding *categories[13]*.

> ***Disaggregation dimensions*** *– the characteristics by which data is to be disaggregated*

> ***Disaggregation categories*** *– the different characteristics under a disaggregation dimension*

1.2.1   Disaggregation Dimensions

The initial set of disaggregation dimensions defined by the IAEG-SDGs classifies data disaggregation dimensions into three categories, as follows:

---

[13] UNSC. 2019. Data Disaggregation and SDG Indicators: Policy Priorities and Current and Future Disaggregation Plans. https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3a-Data-Disaggregation-E.pdf

Category 1- dimensions mentioned in the goals or targets (also referred to as the minimum set of data disaggregation)

Category 2- dimensions for which data are currently available (includes other dimensions in addition to the minimum set)

Category 3- dimensions for which the international custodian agencies may introduce in the future (includes other dimensions in addition to the minimum set)

Table 1-1 presents all of the disaggregation dimensions in the minimum set and the relevant SDG indicators for which data are currently available and will be available in the future[14].

Table 1.1: Indicators that disaggregate data or have plans to disaggregate data for the minimum data disaggregation dimensions[15]

Dimension	Current Available	Future Available	Note
Disaggregation Dimensions Included in Chapeau			
1- Age	1.1.1, 1.3.1,1.4.1 3.2.1, 3.3.1,3.3.2, 3.3.4, 3.4.2,3.5.2, 3.7.1, 3.7.2, 3.a.1 5.2.1, 5.3.1,5.4.1, 5.5.1 8.5.2 16.2.2	1.2.1, 1.2.2,1.5.1, 2.1.1, 2.1.2,2.3.1, 2.3.2 3.1.1,3.1.2, 3.2.2, 3.3.5, 3.6.1, 3.8.2, 3.9.1,3.9.2, 3.9.3 4.2.1,4.3.1, 4.4.1, 4.6.1 5.2.2,5.3.2, 5.6.1,5.6.2, 5.b.1 8.5.1,8.7.1,8.10.2 9.5.2 10.1.1, 10.2.1, 10.7.1, 11.1.1,11.2.1, 11.5.1, 11.7.1,11.7.2 13.1.1, 13.b.1 16.1.1, 16.1.2,16.1.3, 16.1.4, 16.2.1,16.2.3, 16.3.2, 16.5.1,16.6.2, 16.7.1, 16.7.2,16.9.1, 16.10.1 17.8.1, 17.19.2	1.1.1: Only for Employed 1.4.1, 2.3.1, 2.3.2, 3.8.2: Household Head 3.1.2: Maternal Age
2- Disability Status	1.3.1 8.5.2	1.5.1 5.2.1, 5.2.2 8.5.1 10.2.1 11.1.1,11.2.1, 11.5.1, 11.7.1,11.7.2	

---

[14] Details are provided in the document *Compilation on Data Disaggregation Dimensions and Categories for Global SDG Indicators* available @ https://unstats.un.org/sdgs/files/Annex%201%20-%20Disaggregation%20Compilation.xlsx
[15] Updated metadata on the SDGs Indicators can be found here: https://unstats.un.org/wiki/display/SDGeHandbook/Home. A pdf version can be downloaded from https://unstats.un.org/wiki/download/attachments/34505092/SDGeHandbook-150219.pdf?version=1&modificationDate=1550246674487&api=v2

Dimension	Current Available	Future Available	Note
Disaggregation Dimensions Included in Chapeau			
		13.1.1 16.6.2,16.7.1, 16.7.2 17.19.2	
3- Ethnicity		3.1.1 4.1.1 5.2.1, 5.2.2, 5.3.1,5.3.2 11.1.1, 11.2.1 16.1.3, 16.3.1	
4- Geographica l Location	1.4.1 2.3.1,2.3.2 4.1.1 5.4.1 6.1.1,6.2.1 7.1.1 11.6.1,11.6.2 15.4.2 17.19.2	1.1.1, 1.2.1,1.5.1, 1.5.2 2.1.1,2.1.2,2.4.1 3.2.1,3.2.2,3.7.1, 3.7.2, 3.8.1,3.8.2, 3.b.1 4.2.1,4.2.2, 4.3.1, 4.4.1,4.6.1 5.2.1,5.2.2,5.3.1, 5.3.2, 5.6.1, 5.b.1 6.4.1,6.6.1 7.1.2 8.5.2, 8.10.2 9.3.2 10.1.1, 10.2.1 11.1.1, 11.2.1,11.3.1, 11.5.1, 11.5.2,11.7.1 13.1.1 16.2.1,16.2.3,16.5.2, 16.6.2,16.7.2,16.9.1 17.8.1	
5-Income (Wealth)	1.3.1, 1.4.1 3.1.1 10.1.1	1.5.1 3.1.2, 3.2.1,3.2.2, 3.6.1, 3.8.1,3.8.2 4.2.1, 4.2.2,4.3.1, 4.4.1, 4.6.1 5.2.1, 5.2.2, 5.3.1, 5.3.2, 5.6.1 7.1.1 8.10.2 11.1.1, 11.2.1, 11.3.1, 11.5.1, 11.5.2 13.1.1 16.1.3, 16.2.1,16.2.3, 16.5.1, 16.9.1 17.10.1, 17.11.1, 17.19.2	

Dimension	Current Available	Future Available	Note
Disaggregation Dimensions Included in Chapeau			
6-Migrant Status	4.1.1, 4.6.1 8.8.1	8.8.2 10.7.1 11.1.1,11.2.1 16.3.1	4.1.1, 4.6.1, 8.8.1, 8.8.2: Migrant / Non-migrant 10.7.1: Type of Migration Process (Documented / Undocumented)
7-Race		11.1.1, 11.2.1	
8-Sex	1.1.1, 1.4.1 3.2.1,3.3.1, 3.3.2, 3.4.1, 3.4.2, 3.5.2, 3.a.1 4.1.1, 4.2.2, 4.3.1, 4.4.1, 4.c.1 5.4.1,5.6.2, 5.b.1 8.3.1,8.5.1, 8.5.2, 8.6.1, 8.7.1, 8.8.1 16.2.2,16.2.3,16.10.1	1.2.1, 1.2.2, 1.3.11.4.2, 1.5.1, 1.b.1 2.1.1, 2.1.2, 2.3.1, 2.3.2 3.2.2, 3.3.4,3.6.1, 3.8.2, 3.9.1, 3.9.2, 3.9.3 4.2.1,4.6.1 5.a.1 7.1.1 8.8.2, 8.10.2 9.2.2,9.3.2, 9.5.2 10.1.1, 10.2.1, 10.7.1 11.2.1,11.5.1, 11.7.1, 11.7.2 13.1.1, 13.b.1 16.1.1,16.1.2, 16.1.3, 16.1.4,16.2.1, 16.3.1, 16.3.2,16.5.1, 16.5.2, 16.6.2,16.7.1, 16.7.2, 16.9.1 17.8.1, 17.19.2	1.1.1: Only for Employed 1.4.1, 2.3.1, 2.3.2, 3.8.2: Household Head 9.3.2, 16.5.2: Manager, ownership

Detailed updated information on this as well as the full set of dimensions can be found on the IAEG-SDG website at https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/[16]. The information includes over 200 other dimensions, presented by goal. These include, for example, education level, employment status, socio-economic status and occupation.

1.2.2 Disaggregation Categories

Recommendations on categories for the various dimensions are still works-in-progress. However, the existing compilations show that a given dimension requires different categories depending on the goal or target and the associated indicator. Two examples are discussed below:

---

[16] In particular, refer to the Excel files https://unstats.un.org/sdgs/files/Annex%201%20-%20Disaggregation%20Compilation.xlsx and https://unstats.un.org/sdgs/files/Annex%202%20-%20Disaggregation%20Availability.xlsx

Indicator	Minimum Required Disaggregation Dimensions	Minimum Required Disaggregation Dimension Available in Global SDG Database (Yes/No)	Disaggregation Category of Minimum Required Disaggregation Dimension
1.3.1 Proportion of population covered by social protection floors/systems, *by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable*	1. Sex 2. Age 3. Employment Status 4. Disability 5. Pregnancy 6. Work-Injury Victims 7. Income	1. No 2. Yes 3. Yes 4. Yes 5. Yes 6. Yes 7. Yes	1. Male / Female 2. Children/Above Retirement Age 3. Employed/Unemployed 4. People with Severe Disabilities 5. Mothers with Newborns 6. Work-injury Victims 7. Lowest Income Quintile
1.4.1 Proportion of population living in households with access to basic services	1. Sex of HH head 2. Age of HH head 3. Employment Status of HH head 4. Geographic Location (Urban/rural) 5. Household Incomes	1. Yes (HH head) 2. Yes (HH head) 3. Yes (HH head) 4. Yes 5. Yes	1. Male / Female 2. 5-year Age groups from 15+ 3. Employed/Unemployed 4. Urban/rural/city 5. Lowest Income Quintile

For the indicators 1.3.1 and 1.4.1, the common minimum required disaggregation dimensions are sex, age and employment status. However, 1.4.1 refers to the sex, age and employment status of the head of household. While "sex" (Male/Female) and "employment status" (employed/unemployed) have the same categories, the categories of "age" differ. This is because, the subgroups in the population (the dimensions) identified in indicator 1.3.1 include children and older persons.

1.2.3 Overview of existing standards

The IAEG-SDGs has compiled existing standards, thoughts, and ideas on data disaggregation for the minimum set of data disaggregation dimensions. The compilation, which is accessible as a 'living document'[17] provides information for each dimension on:

---

[17] Refer to https://unstats.un.org/sdgs/files/Overview%20of%20Standards%20for%20Data%20Disaggregation.pdf

- Different categories already in use

- Information and evaluation of the categories

- Corresponding categories used in the Global Indicator Framework

- Existing global standards

- Existing regional standards

Information drawn from this compilation for the dimensions of age and ethnicity are shown in Table 1.2 below.

Table 1.2. Standards for categories of Age and Ethnicity

Dimensions	Different categories already in use	Information/ Evaluation	Categories used in the Global Indicator Framework	Existing global standards	Existing regional standards
Age	• Date of Birth • Age groups • 1-year-age-groups	• Use of different age groups in national and international data • Differing age groups demanded in indicator or target	Differing age groups: Commonly used categories • 15-49 • <15 • 15-49 • >15 • 15-65 • <5	UN definition of age groups: • Infants: 0-5 years • Children: 0-15 years • Youth: 5-24 years, (UN Youth) • Adults 15 years and older • Older Persons: 60 years and older	**EU-Standard**: differing age groups Often 10-year intervals are used; e.g. in the EU SDI database
Ethnicity	• Ethnic ancestry or origin • Ethnic identity • Cultural origins • Race • Minority status • Tribe • Language • Religion • Ethnic Self-identification • Recognised (national)	• UN Concepts and definitions: "[…] By the nature of this topic, these categories and their definitions will vary widely from country to country; therefore, no internationally accepted criteria are possible." • UN Standards and Methods: "Ethnicity is multidimensional and is more a process than a static concept, and so	Data is not disaggregated by ethnicity	No international standard possible due to varying national circumstances	EU-Standard: SDG data is not disaggregated by ethnicity Country/type of citizenship

Dimensions	Different categories already in use	Information/ Evaluation	Categories used in the Global Indicator Framework	Existing global standards	Existing regional standards
	minorities	ethnic classification should be treated with movable boundaries." Caution: different connotation of origin and tribe Disaggregation categories could offend certain population groups			

1.3 Developing a Strategy for Getting Disaggregated Data

The task of producing the levels of disaggregation of SDG monitoring indicators needed to better understand the situation of the most vulnerable and marginalized people is a difficult one. National statistical systems need to prepare a strategy to improve the quality, quantity, and availability of disaggregated data and statistics. That strategy should include financing to support capacity to produce and use these.

The Inclusive Data Charter (IDC)[18] is an initiative of the Global Partnership for Sustainable Development Data (GPSDD) to mobilize political commitments and meaningful actions to help countries advance disaggregated data. The strategy and preparation of action plans are guided by the five principles described in Figure 1-2.

In general, action plans are monitored on a voluntary basis.[19]

Figure 1.2: Inclusive Data Charter Principles

Principle One - All populations must be included in the data
We can only achieve the "leave no one behind" goal by empowering the furthest behind. This means ensuring their voices are heard and their experiences are represented through data and analytics. We need to acknowledge all people, make them visible in the data to understand their lives, and inc-lude them in the development process.

Principle Two - All data should, wherever possible, be disaggregated in order to accurately describe all populations
We recognize that data should be disaggregated by sex, age, geographic location, and disability status and, where possible, by income, race, ethnicity, migratory status, and other characteristics relevant in national contexts.

Principle Three - Data should be drawn from all available sources
We recognize the need to make high-quality, timely data from official and non-official sources accessible, and that these should include new data sources, where consistent with internationally accepted statistical standards.

Principle Four - Those responsible for the collection of data and production of statistics must be accountable
We will balance the principles of transparency - maximizing the availability of disaggregated data – confidentiality, and privacy to ensure personal data is not abused, misused, or putting anyone at risk of identification or discrimination, in accordance with national laws and the Fundamental Principles of Official Statistics.

Principle Five - Human and technical capacity to collect, analyze, and use disaggregated data must be improved, including through adequate and sustainable financing
We recognize that collecting and analyzing disaggregated data needs specific skills and these must be built. We recognize the need to finance data collection, analysis, and use appropriately and sustainably so that high-quality data can be collected and used by governments as well as by businesses, civil society, and citizens.

Source: Global Partnership for Sustainable Development Data. 2018. IDC Vision and Principles.

Statistics Sweden (2020a, 2020b) prepared its first report on the situation of those who are furthest left behind in Swedish society through its annual statistical review process that focused on the LNOB principle. This initial review identified sources, breakdowns and specific indicators

---

[18] Global Partnership for Sustainable Development Data. 2018. IDC Vision and Principles. Additional information are available here: https://www.data4sdgs.org/sites/default/files/2018-08/IDC_onepager_Final.pdf

[19] An example of an action plan and an annual monitoring report for the Philippine Statistics Authority can be found here: https://www.data4sdgs.org/sites/default/files/2018-07/PSA%20IDC%20Action%20Plan.pdf and https://www.data4sdgs.org/sites/default/files/2019-07/PSA%20-%20IDC%20annual%20monitoring%20form%202019.pdf, respectively.

that would be incorporated in the future regular statistical reviews. It adopted a pragmatic approach: Use what we have, Use what we know, Learn and develop.

References

Global Partnership for Sustainable Development Data. 2019. IDC Vision and Principles. https://www.data4sdgs.org/sites/default/files/2018-08/IDC_onepager_Final.pdf

United Nations. 2019. *Data Disaggregation and SDG Indicators: Policy Priorities and Current and Future Disaggregation Plans.* https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3a-Data-Disaggregation-E.pdf

United Nations. 2018. Overview of Standards for Data Disaggregation @ https://unstats.un.org/sdgs/files/Overview%20of%20Standards%20for%20Data%20Disaggregation.pdf

United Nations. 2019. *Compilation on Data Disaggregation Dimensions and Categories for Global SDG Indicators.* https://unstats.un.org/sdgs/files/Annex%201%20-%20Disaggregation%20Compilation.xlsx

United Nations. 2019. *Compilation on Data Disaggregation Dimensions and Categories for Global SDG Indicators.* Annex 1- Compilation. https://unstats.un.org/sdgs/files/Annex%201%20-%20Disaggregation%20Compilation.xlsx

United Nations. 2019. *Compilation on Data Disaggregation Dimensions and Categories for Global SDG Indicators.* Annex 2- Availability. https://unstats.un.org/sdgs/files/Annex%202%20-%20Disaggregation%20Availability.xlsx

UNDP. 2018. What does it mean to leave no one behind? A UNDP discussion paper and framework for implementation. https://www.undp.org/content/undp/en/home/librarypage/poverty-reduction/what-does-it-mean-to-leave-no-one-behind-.html

Statistics Sweden. 2020a. *Statistical review focusing on the principle of leaving no one behind.* Paper prepared for the UNECE Expert Meeting on Statistics for SDGs, 22-27 April 2020. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.32/2020/mtg1/W_1_2_ENG_Sweden-Statistical_review_focusing_on_the_principle_of_Leaving_no_one_behind.pdf

Statistics Sweden. 2020b. *Annual statistical review with a focus on LNOB.* Presentation at the 11the IAEG-SDG Meeting, November 2020. https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-11/13b.%20Sweden-first%20country%20report%20with%20a%20focus%20on%20vulnerable%20populations_Sweden.pdf

Chapter 2. Integrating Policy Demands on Inclusive Development with Data

Overview

The "leave no one behind" (LNOB) principle requires that all vulnerable groups are identified and targeted by policies. Striking a balance between policy formulation and demand for data calls for active involvement of data producers in policy processes. A human rights-based approach to data allows the collection and publication of disaggregated data. It is toward this end that the IAEG-SDGs mapped the policy priority targets and indicators to provide countries with concrete suggestions of areas for data disaggregation. This chapter provides an overview of three examples of tools that can facilitate the linking of policies and data—EPIC, ADAPT and StatAct.

2.1   The policy-data nexus for leaving no one behind

With the 2030 Agenda for Sustainable Development, data-driven policy formulation has taken on greater significance. To "leave no one behind", statistical processes must first ensure that everyone is counted. LNOB requires that all target groups—the vulnerable groups-- that are likely to be left behind be identified and acknowledged by the policy and included in the data production and reporting process. Vulnerability is multidimensional and policy and data need to apply various lenses (social, economic, and environmental) in an integrated manner to correctly zero in on the target vulnerable groups. What happens if this is not done, as is often the case? Aura's case described in Box 2.1 below illustrates the consequences.

Box 2.1: *Aura's Case*

Aura is an 85-year-old woman living in South Intopia, the most polluted part of the country where the economy is dominated by the mining industry. She suffers from acute asthma and earns an equivalent of 2 USD per day from her husband's pension plan. In recent years, Aura has struggled with her medical expenses and finally decided to sign up for a programme that distributes cash to poor families. The *Ministry of Housing and Poverty* has identified Aura ineligible for the programme as she earns above the minimum daily income of 1.90 USD. She explained her exceptional condition to the officer and was told that she is "***economically***" not poor, so her problem had to be taken to the *Ministry of Health* where she was told last year that her *disease was not on the list* of the subsidized medical coverage plan. She was also advised by her doctor to move out of the polluted area. Aura had sought support from the *Ministry of "**Social**" Welfare* to relocate her house, but her justification was "***environmental***" which falls outside the mandate of this ministry and furthermore she was not defined as poor! So, Aura should take her issue back to the Ministry of Housing and Poverty, which means that she would *repeat a futile cycle*.

Aura cannot find any place in the policy priorities. She is not "left behind" according to data and not prioritized by any of the policy programmes. Why do the data not show her as a vulnerable member of the society? Why has no policy identified her as target/vulnerable citizen? And, who should fix the problem-- data producer or policy maker?

One increasingly necessary conclusion is, data should be made available in a format that allows the identification, as well as examination, of multiple and intersectional characteristics that define many of those who are left behind.

Source: Bidarbakht Nia. 2018. Why, What and How of Policy – Data Integration.
https://www.unescap.org/sites/default/files/S3%20-%20Policy%20Statistics_Data%20integration.pdf.

Approaches to policy frameworks that integrate the principle of LNOB have included using the principle as a guide for specific targeted interventions or as a guide to an overall development strategy. Regardless of what is adopted, trade-offs are expected in policy formulation.

A preferential focus on ensuring that no one is left behind can be very costly, as well as politically challenging, and may come at the cost of promoting the efficient use of scarce resources. Examples of the resulting tensions-- should infrastructure be focused on areas lacking progress, even if this means that economic development of the entire country slows down as a result? Similarly-- should one focus on improving the lives of those left behind where they currently live which may be costly, or increase their opportunities to move to more economically dynamic regions? Should a focus on improving the lives of those left behind take precedence over other goals, such as environmental sustainability? Clearly, these are difficult issues that require detailed country-specific analysis and policy responses which are often difficult, as pointed out in Box 2.2.

Box 2.2: Policy-data and data-policy—A vicious cycle

It is often a case of the chicken or the egg dilemma when it comes to policy-data integration.

The lack of data is blamed for the absence of appropriate policies due to insufficient evidence, whilst the lack of demand is seen as the main challenge for producing relevant data.

The root cause analysis begins with "lack of demand" or "lack of supply" for data, depending on who the analyst is - the data producer or the policymaker.

Moreover, data producers don't actively engage in policy discussions to see where the evidence for policymaking is missing, and policymakers often fail to specify what data is needed for monitoring.

This has created a vicious cycle, where the give-and-take of policymakers and producers of official statistics recurs with little success.
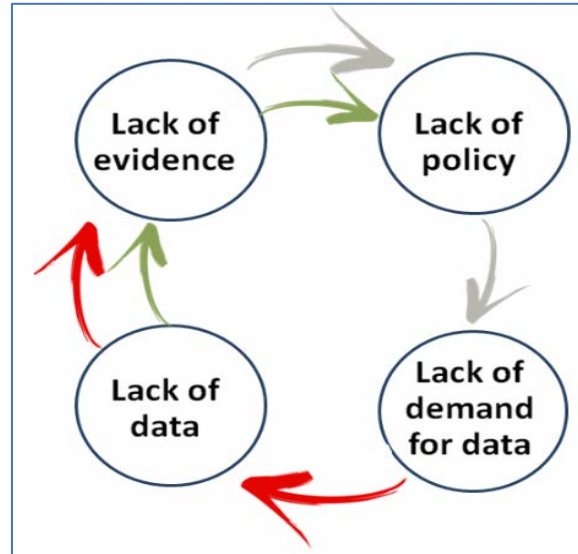
Source: Bidarbakht Nia (2019): Connecting policymakers and data producers https://www.unescap.org/blog/connecting-policymakers-and-data-producers.

Policy processes involve different stages, from agenda-setting, formulation and implementation to monitoring and evaluation. Once an issue requiring a policy solution has been identified, the

process of policy development includes how an issue is framed by various stakeholders, which issues get prioritized in the policy agenda for a given period, and how the policy is formulated. Decisions at these various stages of policy development, including adjustments to policies, should be better informed by available data and statistics. Policies should also include rational analysis of what works and what does not.

Policies that are based on systematic evidence are widely viewed to produce better outcomes than those not informed by data and statistics. The policy-data nexus, illustrated in Figure 2.1 argues that the absence of appropriate policies can be attributed to data gaps, but lack of demand for data is also a primary challenge for producing relevant data. Whether lack of data is due to a "lack of demand" or "lack of supply", it is important for data producers to be actively part of policy discussions in order to share perspectives.

Figure 2.1: Policy-data and data-policy

Source: *Bidarbakht Nia (2019): Connecting policymakers and data producers*
https://www.unescap.org/blog/connecting-policymakers-and-data-producers.

2.2   Disaggregated data—The human-rights based lens

The 2030 Agenda for Sustainable Development is unequivocally anchored in human rights with the 17 SDGs, 169 targets, and 231 indicators meant to "realize the human rights of all". It is explicitly "grounded in the United Nations Charter, the Universal Declaration of Human Rights, international human rights treaties" and other instruments, including the Declaration on the Right to Development.

Data collection and disaggregation that allow for comparison of population groups or characteristics are central to a human rights-based approach to data (HRBAD)[20] and forms part of States' human rights obligations—that is, to collect and publish data disaggregated by grounds of discrimination recognized in international human rights law.

The HRBAD guidance note points out:

> *"As devising disaggregation of indicators (or not) is not a norm or value neutral exercise, and the risks associated with this operation for the protection of the rights of data subjects cannot be denied, an HRBAD has much to offer in this context. As outlined in this note, an HRBAD helps bring together relevant data stakeholders and develop communities of practice that improve the quality, relevance and use of data and statistics consistently with international human rights norms and principles."*

Components of the Human-Rights Approach to Data

---

[20] UN OHCHR. 2018. Guidance note to data collection and disaggregation.
https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf

This approach identifies a preliminary set of principles, recommendations and good practices under each of six interrelated components (refer to Figure 2.2), briefly described below:

Participation of relevant population groups in data collection exercises, including planning, data collection, dissemination and analysis of data

Data disaggregation allows data users to compare population groups, and to understand the situations of specific groups.

Self-identification. For the purposes of data collection, populations of interest should be self-defining. Individuals should have the option to disclose, or withhold, information about their personal characteristics

Transparency. Data collectors should provide clear, openly accessible information about their operations, including research design and data collection methodology. Data collected by State agencies should be openly accessible to the public

Privacy. *Data disclosed to data collectors should be protected and kept private, and confidentiality of individuals' responses and personal information should be maintained*

Accountability. *Data collectors are accountable for upholding human rights in their operations, and data should be used to hold States and other actors to account on human rights issues*

Figure 2.2: Six components of a human rights-based approach to data

Source: United Nations. 2018. A human rights-based approach to data: Guidance Notes to Data Collection and Disaggregation. https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf.

Data disaggregation and the HRBAD principles

Data collection and disaggregation that allow for comparison of population groups are central to an HRBAD and forms part of States' human rights obligations. Disaggregated data can inform on the extent of possible inequality and discrimination.

Key HRBAD principles on the disaggregation of data are:

More detailed data than national averages is key in identifying and understanding inequalities

Data should be disaggregated by key characteristics identified in international human rights law

Collection of data to allow disaggregation may require alternate sampling and data collection approaches

Birth registration is foundational for robust data sets that allow accurate disaggregation

A compilation by OHCHR, *International human rights standards and recommendations relevant to the disaggregation of SDG indicators[21],* lists human rights standards and recommendations for the following population groups or characteristics:

- Gender/Sex

- Age (children, adolescents, older persons)

- Race/colour/ethnicity/national origin/birth

- Indigenous peoples

- Minorities

- People of African descent

- Persons with albinism

- Migrants/internally displace persons/trafficking in persons

- Geographic status

- Health status/HIV/AIDS

- Disability

- Religion, belief and ideology/political affiliation/professional status

- Civil status

- Income/socio-economic status

- Sexual orientation/gender identity/intersex status/sex characteristics

2.3   Vulnerable Groups in the SDGs and Policy Priorities

The global indicator framework for the SDGs and targets as originally developed by the IAEG-SDGs and adopted by the UN General Assembly and its current version reflecting annual refinements[22] identify the populations to be monitored by listing key disaggregation variables in the chapeau, as follows:

> *"Sustainable Development Goal indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics, in accordance with the Fundamental Principles of Official Statistics."*

SDG target 17.18 recognizes the need for more systematic data disaggregation, calling on all parties to—

---

[21] More information are available here: https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-07/Human%20Rights%20Standards%20for%20Data%20Disaggregation%20-%20OHCHR%20-%20Background%20Document.pdf
[22] Refer to: https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202020%20review_Eng.pdf. For all UN languages see: https://unstats.un.org/sdgs/indicators/indicators-list/.

"… increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts."

Toward this end, the IAEG-SDGs initial work on data disaggregation included an extensive consultative exercise that identified policy priorities for 'vulnerable' population groups to be defined by application of the dimensions and categories for disaggregation discussed in Chapter 1. The consultation aimed to provide a better understanding of what the corresponding policy areas are and to provide advice on the dimensions and categories necessary for disaggregation of these priority areas. The policy priority areas, when mapped to SDG targets and indicators, provide countries with concrete suggestions of areas for data disaggregation. The initial results of this work-in-progress are briefly summarized in Table 2-1 below[23].

Table 2.1: Policy priorities for vulnerable populations

Population Group	Policy Priorities
Poor	Inclusive and pro-poor growth Social protection systems and floors Effective governance, including participation and use of available resources
Women and girls	Poverty eradication Food insecurity and health Education Access to economic resources and decent work for all Gendered impacts of climate change
Children	Every child survives and thrives Every child learns Every child is protected from violence and exploitation Every child lives in a safe and clean environment Every child has an equitable chance in life
Older persons	Health Income security Violence, abuse and safety Empowerment and participation as full members of society
International migrants	Provide access to basic services (education and health care) and social protection for migrants Ensure fair recruitment, decent work and labour rights protection for migrants Eliminate all forms of discrimination and reduce violence against migrants Ensure access to information for migrants**.**

---

[23] UNSC. 2019. Data Disaggregation and SDG Indicators: Policy Priorities and Current and Future Disaggregation Plans. https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3a-Data-Disaggregation-E.pdf. Detailed information can be accessed here: https://unstats.un.org/sdgs/files/Policy%20Priorties%20Inputs%20compilation.zip

| Population Group	Policy Priorities
Forcibly displaced persons (refugees and internally displaced persons)	Basic needs and living conditions Livelihoods and economic self-reliance Civil, political and legal rights
Persons with disabilities	Poverty eradication Education Employment Health Accessibility

2.4   Tools for policy-data integration

This section provides overviews of three tools that can facilitate the linking of policies and data—EPIC, ADAPT and StatAct.

2.4.1   Every Policy is Connected (EPIC)

Data disaggregation requirements for the SDGs must ultimately be identified in a process of policy-data integration. For this purpose, the development community has developed various policy-data integration tools to assist governments. ESCAP has developed EPIC (Every Policy Is Connected), a tool for reviewing the data needs for monitoring national policies. The tool is based on a framework for a structured, participatory and principle-based dialogue between policy and data stakeholders.

The framework first defines a policy-data landscape and identifies: (i) gaps (lack of demand, supply, and use of data); (ii) voids (absence of policy and therefore data); and (ii) waste in the production of data and formulation of policy. The relationships among these concepts are depicted in Figure 2.3.

Figure 2.3: Policy-data space
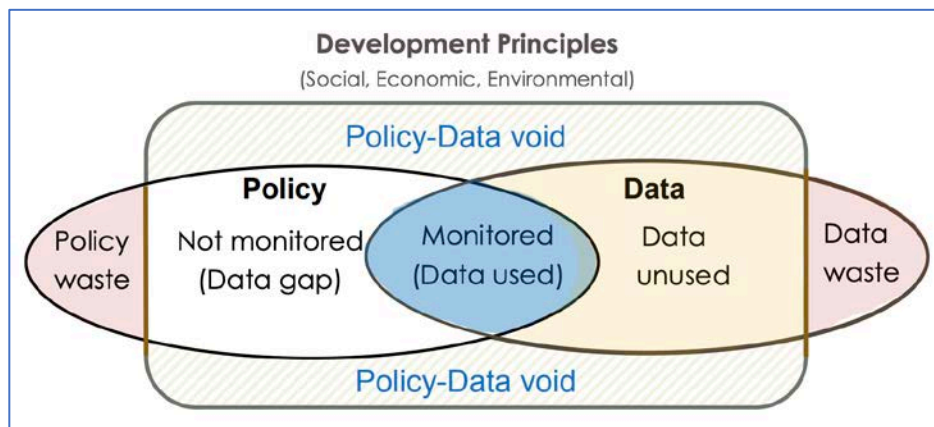
Source: Bidarbakht Nia (2018): Policy-Data Integration: key to achieving the SDGs for all[24]

EPIC reviews all policy documents related to one theme with respect to its coherence with a set of agreed principles and intended beneficiaries/target groups. The framework thus maps out the social, economic, environmental, and institutional dimensions of every issue that requires policy action. It also identifies target groups that will benefit from policy implementation and develops indicators for the key issues identified. EPIC produces a diverse set of outputs that include indicators needed for the current policies, data gaps for monitoring agreed policies, new agreements on policy priorities and inputs to future work for filling the current policy void.

Figure 2.4:

For more details on EPIC refer to: Working Paper Series (SD/WP/09/September 2019): Every Policy Is Connected (EPIC): A generic tool for policy-data integration.https://www.unescap.org/resources/working-paper-series-sdwp09september-2019-every-policy-connected-epic-generic-tool-policy

---

[24] https://www.unescap.org/resources/working-paper-series-sdwp07april-2018-policy-data-integration-key-achieving-sdgs-all
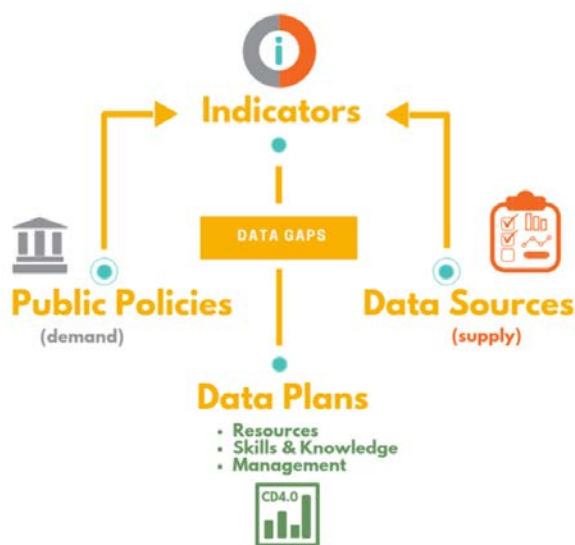
2.4.2 Advanced Data Planning Tool (ADAPT)

PARIS21 developed a complementary policy-data tool called the Advanced Data Planning Tool (ADAPT), which examines the cost of development of specific SDG indicators (refer to Figure 2-5). This is an online tool meant specifically for national SDG managers. ADAPT goes beyond currently available data and begins by setting policy targets. The tool supports the process of indicator development when there is agreement on policy targets. If data are not available, the tool helps estimate the costs of compilation of the SDG indicator/s. In the context of the policy-data nexus for SDG indicators, ADAPT can help by presenting reports (with charts and tables) on: (a) the policy relevance of applicable SDG indicators; (b) availability of applicable SDG indicators; (c) available SDG indicators by prevailing data sources; (d) levels of dependency on external technical and financial assistance for available SDG indicators; and (e) feasibility to compile and produce applicable but unavailable SDG indicators.

ADAPT has a Gender Module that allows the monitoring of the production of data for gender statistics, particularly those linked to relevant global commitments and national policies. The module assists users in tagging indicators as gender-relevant according to national and global priorities. Reports that can be produced when using the module include indicators that are sex disaggregated or have been tagged as gender-relevant across national and global policies, such as the 54 SDG gender-specific indicators identified by UN Women and the 52 Minimum Set of Gender Indicators determined by the Inter-Agency and Expert Group on Gender Statistics (IAEG-GS).

NSOs of Cambodia, Cameroon, Equatorial Guinea, Maldives, Mongolia, Philippines, Rwanda, Tanzania, among others, have made use of ADAPT. The tool was also introduced in 2018 across 20 countries supported by the UNSD and the United Kingdom[25].

Figure 2.5: The advanced data planning tool (ADAPT): Linking data and policies

---

[25] More information are available here: https://www.un.org/development/desa/capacity-development/projects/project/statistical-capacity-for-progress-on-sdgs/

2.4.3 StaTact

Figure 2.6: StaTact

Source: https://statact.unitar.org/en/about-statact

StaTact was developed by the United Nations Institute for Training and Research (UNITAR) and the UN Statistics Division to enable countries to address measurement gaps that impede monitoring national policies and leverage the 2030 Agenda to help resolve problems tactically. It provides an analytical framework and a multi-stakeholder methodology that facilitates commitment and collaboration of a team of national experts from National Statistical Offices, concerned Ministries and other parts of the data community, including non-traditional data sources to jointly design a short-term action plan that addresses institutional impediments to data collection, production and utilization. StaTact aims to provide cost effective solutions for resolving data gaps to address urgent policy need goals within one year. The tactical (quick fix solutions) complements ADAPT which is a longer-term strategic tool.

After having been used by 15 pilot countries, primarily African and Asian LDCs, and one SIDS country in 2018, StaTact moved online in early 2019 that has been used by 13 Small Island Developing States.

More information about StaTact is accessible @ https://statact.unitar.org/en/about-statact.

Resources

Recommended Reading	
	OECD. 2018. Development Cooperation Report 2018: Joining Forces to Leave No One Behind, OECD Publishing, Paris. https://doi.org/10.1787/dcr-2018-en Part I of the report provides evidence of why leave no one behind matters along with data and analysis on what it means to be left behind. Chapters zoom-in on eight critical issues that need to be tackled to achieve the SDGs for all: ending extreme poverty in countries most in need; tackling rising income inequality; addressing fragility; enabling inclusive governance; the imperative of climate action; making progress towards gender equality and women's economic empowerment; including the world's 1.2 billion young people; and ensuring persons with disabilities are no longer left behind.

Part II investigates leaving no one behind in practice. Chapters shed light on the potential impact of more integrated policies, budgets and programmes across sectors and between levels of government in reaching the most vulnerable. Achieving the SDGs for all relies on data and diagnostics that count everyone and are disaggregated by factors like income, sex and gender, geography, age, and disability.

Chapters 2, 3, 5 and 8 are especially instructive— dealing with topics related to Chapters 1 and 2—and segues into Chapter 3.

Recommended Reading	
	UN OHCHR. 2018. Guidance Note to Data Collection and Disaggregation. https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf The components and key principles of human rights-based approach to data (HRBAD) are briefly discussed in this Chapter. This publication provides more detailed discussions and recommendations and good practices.

Reference

Bidarbakht Nia, Arman. 2018. Policy-Data Integration: key to achieving the SDGs for all. ESCAP Statistics Division Working Paper Series #07.
https://www.unescap.org/sites/default/d8files/knowledge-products/SD_Working_Paper_no.7_Apr2018_Policy-Data_Integration.pdf

Bidarbakht Nia, Arman. 2019. Connecting policymakers and data producers. ESCAP Blog.
https://www.unescap.org/blog/connecting-policymakers-and-data-producers#

GPSDD. Data4SDGs Toolbox. *In progress.* A set of tools, methods, and resources to help countries to create and implement their own holistic data roadmaps for sustainable development that address institutional, policy, technical, resource, and capacity issues, among other things. https://www.data4sdgs.org/initiatives/data4sdgs-toolbox (accessed on 14 October 2020)

Klasen, S. and Fleurbaey, M. 2018. Leaving no one behind: Some conceptual and empirical issues. UN Department of Economic & Social Affairs CDP Background Paper No. 44 ST/ESA/2018/CDP/44. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/CDP_BP44_June_2018.pdf (accessed on 23 June 2020)

Partnership in Statistics for Development in the 21st Century (PARIS21). n.d. Advanced Data Planning Tool (ADAPT). Paris, France: PARIS21. https://paris21.org/advanced-data-planning-tool-adapt (accessed on 23 June 2020)

United Nations. 2014. A/RES/68/261- Fundamental Principles of Official Statistics Resolution adopted by the General Assembly on January 29, 2014. New York City, NY, USA: UN. https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf (accessed on 23 June 2020)

United Nations. 2015. The Millennium Development Goals Report 2015. http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf (Accessed on 23 June 2020)

United Nations. 2016. Transforming Our World: The 2030 Agenda for Sustainable Development. https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf (Accessed on 23 June 2020)

United Nations Economic and Social Commission for Asia and the Pacific. 2018. EPIC: a generic tool for policy-data integration. https://www.unescap.org/sites/default/files/EPIC_V1.2_AB15Aug18.pdf (accessed on 23 June 2020)

United Nations Office of the High Commissioner for Human Rights. A Human Rights-Based Approach to Data: Leaving No One Behind in the 2030 Agenda for Sustainable Development. https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf (accessed on 23 June 2020)

United Nations Office of the High Commissioner for Human Rights. Transforming Our World: Human Rights in the 2030 Agenda for Sustainable Development. https://www.ohchr.org/Documents/Issues/MDGs/Post2015/TransformingOurWorld.pdf (accessed on 23 June 2020)

United Nations Office of the High Commissioner for Human Rights. International human rights standards and recommendations relevant to the disaggregation of SDG indicators. https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-07/Human%20Rights%20Standards%20for%20Data%20Disaggregation%20-%20OHCHR%20-%20Background%20Document.pdf (accessed on 23 June 2020)

Chapter 3. Sources of Disaggregated Data: Strengths, Potentials and Limitations

Overview

Before embarking on a disaggregation exercise, there are many issues to consider. This includes specific challenges for different data sources. For example, as large sample sizes are needed for disaggregation it can be expensive to use household surveys. IOM GMDAC conducted a pilot study on disaggregating SDG indicators by migratory status based on harmonized census data for 73 countries. However, census data often takes years to be released before it can be analysed. As a result, NSOs may explore the potential of administrative data sources or census microdata for SDG data.

There are challenges in defining disaggregation categories – some important concepts relevant to disaggregation lack internationally agreed definitions. Further, there are important considerations beyond data collection. For example, there may be a need for specialised awareness raising in NSOs on the importance of disaggregation in the first place, or for tailored approaches to disseminate disaggregated data to policymakers in the right ministries.

This chapter looks at main sources of disaggregated data, illustrating how they have been successfully used, the inherent limitations of the sources, how countries are applying innovations to better tap their potentials for meeting disaggregation requirements for SDG indicators and the current capacity limitations.

3.1 Sources of disaggregated data: Census

Two approaches to data collection that official statistical systems have traditionally relied on are censuses and sample surveys. A census is generally a study of every unit (everyone or everything) in a specified population. A census is often described as a complete enumeration or a complete count of population units. In a sample survey, on the other hand, data is collected from a part (sample) of the total population. Both approaches provide data that are used to draw conclusions about the whole population. This section looks at censuses as sources of disaggregated data. Main types of censuses differentiated by the units of observation are population and housing censuses, business and industry census, and the agriculture census.

Censuses provide disaggregated data in several ways.

A census of population and housing can provide data to calculate directly indicators on the topics it covers. Typically, censuses cover the following disaggregation dimensions: sex/gender, age, ethnicity/race, disability status, migrant status, education level and geographic location (e.g., administrative or subnational areas; urban/rural).

As a sample frame for surveys, providing the reference population for the special/vulnerable population groups. For example, censuses where identification of disability status of persons is part of the data items that can be used to construct sampling frames for disability surveys.

As inputs to small area estimation modeling to produce disaggregated data.

Geospatial data analysis.

Potentials

Innovations to censuses and surveys to address SDG needs are exemplified in the FAO initiatives described in Box 3.1.

Box 3.1: What's new in data collection for SDGs in agricultural censuses and surveys?

The FAO World Programme for the Census of Agriculture (WCA) provides support and guidance to countries to carry out national agricultural censuses. Data collected provides a snapshot of the state of a country's agricultural sector - from size of holdings, land tenure, land use, area harvested, irrigation, livestock, labour and other agricultural inputs. This information is vital in agricultural planning and policymaking, research and development and monitoring the impact of agriculture on the environment.

For this 10 year-round, FAO has developed a new set of guidelines that involve a number of developments taking into account the changing nature of data use and collection. New data collection technologies such as computer-assisted personal interviewing (CAPI) have been included to improve data quality and drastically reduce the time lag between data collection and data analysis. The guidelines also emphasize the use of interactive outputs and web-based data (e.g., tables, graphs, maps), and access to anonymized micro-data, as new prospects for census dissemination. Additional data domains have been included/proposed, such as fisheries and GHG emissions to aid monitoring the achievement of the Sustainable Development Goals (SDGs). Among the new methodologies proposed, there is also a renewed approach for assessing food security by reviewing the severity of food insecurity as experienced by individuals in the population. For this round, the WCA has also enhanced the process for measuring the distribution of managerial decisions and initiated the identification of ownership within the household for the collection of sex-disaggregated data (FAO 2015). Additionally, the FAO and World Bank Global Strategy; the Evidence on Data and Gender Equality (EDGE) project, which has collaborated with FAO on clear recommendations for agricultural censuses to include sex disaggregated data on land ownership; and an FAO report and guide on integrating gender in agricultural surveys are recent global initiatives that aim to resolve the lack of sex-disaggregated data in agricultural censuses and survey. Major gains have been made among household surveys after the recent adoption of the SDGs and efforts of international agencies to address gender data gaps in agriculture through enhancements of household survey directed at the household and agricultural landholding levels. The World Bank's Living Standards Measurement Study – Integrated Surveys on Agriculture, for instance, have been used frequently in recent research of men's and women's outcomes in agriculture in Sub-Saharan Africa. The Global Environmental Facility's Integrated Approach Pilot is another project-specific household survey, which is examining how to further mainstream sex-disaggregated data in survey modules on agricultural productivity and climate change (FAO 2019).

A UNFPA review[26] concludes that population and housing censuses can potentially provide data for 14 out of 17 goals with population related elements present in approximately 45% (105 of 232)

---

[26] Source: PowerPoint presentation @ Workshop xxx

of the SDG indicators. There are 19 census indicators, i.e., SDG indicators that can be calculated from a census exclusively with the census identified as the data source.

A new methodology to generate spatially disaggregated population estimates through a hybrid census is described in Box 3.2.

Box 3.2: New Methodology: a hybrid census to generate spatially disaggregated population estimates[27]

Within the 2030 Agenda, the growing requirement for spatially disaggregated population data has triggered the exploration of new data sources at different geographical scales and time periods, especially in highly stressed countries and countries without a recent population census. Advances in the availability of detailed satellite imagery, geo-positioning tools for field surveys, statistical methods and computational power are enabling the development and application of approaches that can estimate population distributions at fine spatial scales across entire countries, in cases where population and housing census data cannot be conducted.

The goal of a hybrid census is to produce population estimates for small areas or uniform, detailed grids in the absence of a traditional national census. Hybrid censuses rely on complete counts of population within small, defined areas, through 'micro-census surveys', selected across an area of interest, and collected relatively rapidly and at a fraction of the cost of a full national census. Statistical models are then used to link these micro-census data to spatial data with full coverage over the regions of interest to predict population numbers in the unsampled locations. By aggregating these high-resolution predictions, population totals can also be produced for administrative units or for the national level if required.

This approach can never replace the rich production of data on the individual, family, household or community generated by a traditional population and housing census. However, where a traditional census cannot be fully executed in all locations of a given country due to insecurity or other concerns, then this hybrid approach could produce population estimates for small areas or uniform, detailed grids in the absence of traditional census data.

Limitations

Censuses are conducted only every 5 or 10 years. As direct sources of data, these would provide at most only two data points for the census indicators. However, census data have longer-term value as bases for population projections and sampling frames; when integrated with other data such as surveys and administrative data, these data are invaluable.

3.2 Sources of disaggregated data: Sample Surveys

A total of 80 indicators, distributed across 13 Goals, are being or could be sourced from household surveys. Around 70 percent of the indicators calculated from household survey data have a requirement of at least one level of disaggregation, supporting the overall goal of "Leaving No One Behind". 37 out of 80 indicators have age as a minimum disaggregation dimension, followed by sex (32), disability (14), and income (8). Household surveys offer the potential to help achieve the minimum disaggregation requirements of the Global SDG Indicator framework since they cover a wide range of topics by design.

Some of the major internationally comparable household surveys include (i) Macro International's Demographic Health Survey (DHS), funded by the United States Agency for International Development (USAID), (ii) the Multiple Indicator Cluster Survey (MICS) overseen by the UN

---

[27] Source: UNFPA. 2017. https://www.unfpa.org/sites/default/files/resource-pdf/Hybrid_Census_Brief_v9.pdf

Children's Fund (UNICEF), and the Living Standards Measurement Study (LSMS) Surveys, which is supported by the World Bank. These household surveys are usually conducted every 3 to 5 years.

Potential

Household survey data provide numerous disaggregation possibilities, of which many are not yet fully exploited. Household surveys offer flexibility in the range of disaggregation dimensions that are covered through sampling approaches and the questions asked. Although household surveys are not well suited to generate spatial disaggregates at lower administrative levels, basic disaggregates such as place of residence, sex, wealth status, and age are usually available. However, innovative approaches such as combinations of these (e.g. urban-poor, young women) or the ability to generate indicators for specific subgroups across many dimensions (e.g. multidimensional poverty) lend more analytical power to the data.

Apart from basic disaggregates, household survey data provide numerous other disaggregates and outcomes that other data systems do not have; household surveys can expand the breadth of SDG indicators covered with ease, compared to administrative systems, which collect small numbers of indicators. Despite this, countries and survey programs do not typically exploit the full potential of microdata. Future work must ensure that adequate further analysis frameworks and tools are developed to better understand the data collected.

Limitations

(a) Sample sizes and reliability

A survey which is intended to compile a (socioeconomic) development indicator typically comprises a sample size that is large enough to provide nationally representative estimates. The sample size generally also provides estimates that fall within tolerable levels of reliability when further disaggregating indicators by broad domains such as states, regions, or provinces. However, the high cost of conducting surveys, prevents a data collection agency to increase the sample size that will afford reliable estimates at more granular levels, such as municipalities and villages.

To substantiate this argument, consider a hypothetical population comprising of 20 million households across 17 regions and 85 provinces. Furthermore, suppose the objective is to estimate a parameter of interest (expressed as a ratio) with a margin of error of 0.05 and a 95% confidence level based on different levels of disaggregation. If one employs simple random sampling, it requires to collect data from roughly 6800 households to be able to afford a nationally representative estimate that meets the pre-specified accuracy levels. If, however, the objective is to provide regionally representative estimates, wherein each estimator meets the pre-specified accuracy thresholds, one needs to sample roughly 115 thousand households. The required sample size further increases to 575 thousand households to afford provincially representative estimates.

Figure 3.1: Recommended Sample Size for Different Levels of Geographic Disaggregation
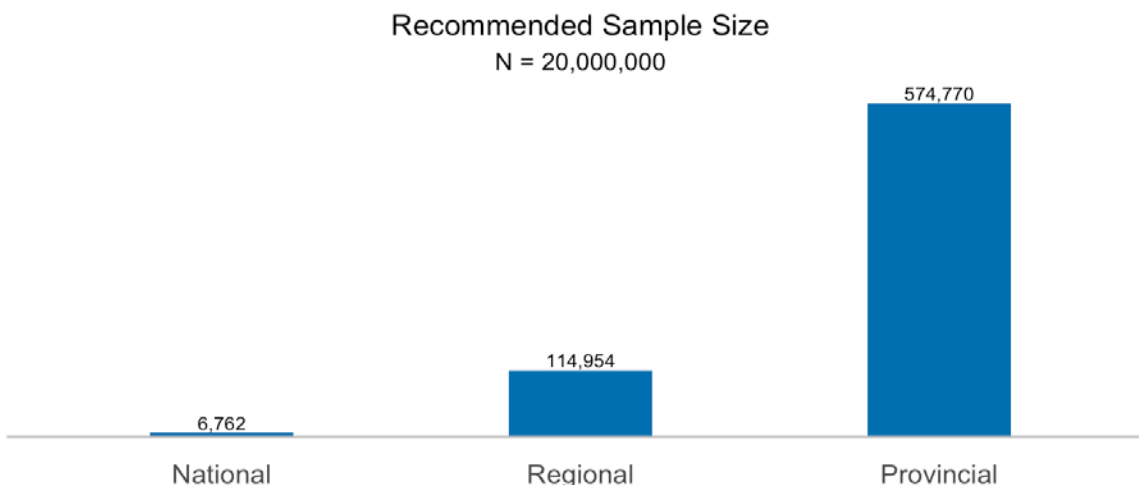
Source: ADB (2020). Introduction to Small Area Estimation Using R Software A Practical Guide for Staff of National Statistics Offices.. https://www.adb.org/publications/small-area-estimation-guide-national-statistics-offices.

This illustration highlights that to afford more granular survey estimates, increasing the survey's sample size is ideal as long as financial and technical resources are not an issue. However, this is not usually true for many national statistical offices or data collection agencies that conduct such surveys. Often, they are constrained with limited budget and resources and this affects collection of data on migrants, the elderly or persons with disabilities.

Sample surveys, which have been typically designed to yield reliable estimates up to national, or regional levels given the costs of data collection, will need increased sample sizes to allow the generation of reliable data for specific sectors of the population, such as migrants, the elderly or disabled. Larger sample sizes require additional financial costs.

(b) Bias due to underreporting on sensitive topics

The sensitivity of some SDGs indicators such as indicators pertaining to violence against women and children (VAWC), or maternal death, could lead family members to under-report these, particularly in household surveys. Further, both maternal mortality and VAWC may be relatively rare. A household survey would thus hardly find respondents that have experienced VAWC or a maternal death (or having a sibling who died of pregnancy-related causes) in their household. The margin of error attached to these estimates of the proportion of households that have experienced VAWC and maternal mortality rate is very large.

(c) Omissions

In some cases, disaggregated data may not be available from traditional data sources because requisite basic data have not been collected. Several household surveys, for instance, omit by design some persons in the population: persons that are "homeless", or who sleep in their work places which are not enumerated as dwellings; those who are in institutions, including refugee camps; and those from nomadic populations. In addition, household surveys typically under-cover those who reside in areas with security risk or that are remotely accessible. The customary length

of a sample survey and the need to maintain confidentiality for the collection of culturally-sensitive data (e.g. ethnicity data) have also constrained production of disaggregated data.

Adjustments in survey designs should be made, or special surveys targeting these omitted populations or undercounted groups in traditional surveys need to be carried out since these population groups are likely marginalized, aside from being invisible from current measurement systems.

3.3 Sources of disaggregated data: Administrative reporting systems

Administrative reporting systems data come from registers of records such as hospital or school records. These are cheaper to collect than survey data, but at present, coverage in many countries is incomplete and data quality can be problematic as data are not usually designed for statistical purposes.

An important type of administrative data is from the civil registration system – a machinery set up in a country for the legal recording of vital events (such as births, marriages and deaths) of the population on a continuous basis.

Potential

Decades ago, NSOs and other data producers in government generated statistics only from censuses and sample surveys, which were designed specifically for statistical purposes. Data sources subsequently expanded to administrative reporting systems as NSSs came under pressure to improve the efficiency of the statistical production process by saving on costs and staff resources. At the same time, there was growing demand to reduce the burden placed upon respondents in sample surveys and censuses. While administrative-based data are rarely direct substitutes for data collected from censuses and sample surveys, it offers possibilities for reducing the cost in statistics generation and the burden on respondents.

Limitations

Data on maternal mortality rates can be alternatively and ideally sourced from vital registration systems, but, across many countries, these data systems do not have full coverage, especially on deaths. While there may be incentives to have births registered, deaths can be seriously underreported. Information on causes of death as collected from death registration systems or even medical certifications may also be misreported.

3.4 Sources of disaggregated data: Small-area estimation (SAE)

Strengths

Sample surveys are designed to produce reliable statistics at a given level of aggregation defined by sampling domains, generally geographical or socio-economic subgroups of the population. This is achieved by, among other design considerations, determining the sample size required to produce reliable estimates for the specified aggregation level. However, there is often a subsequent demand for estimates at a more disaggregated level. Producing statistical estimates for more disaggregated levels than were originally planned from survey data without increasing

the survey sample size (and hence, cost of data collection), is the objective of small area estimation (SAE).

> A *small area* may refer to a *geographic area*, which is more disaggregated than the geographic domain for which the survey is originally designed to provide reliable estimates. Going back to the example on the hypothetical population of 220 million households distributed among 85 provinces within 17 regions, regions and provinces could be both considered as small areas.
>
> In addition, the term *small area* may refer to *socio-economic subgroups*, such as age-sex-ethnic group of persons, and the vulnerable groups targeted by Agenda 2030.

SAE is a class of statistical techniques which are anchored on the notion of combining multiple data sources and capitalize on the strengths of each source. Consider, for example, the strengths of a survey and those of a census. Compared to a census, a survey is a more flexible data collection vehicle in terms of the range of information it can collect. In contrast, a census typically collects less detailed information. But as discussed earlier, there is also a flipside when conducting surveys as survey estimates are subject to margins of error which could be inflated if the sample size is too small. SAE techniques include ones that combine a survey and a census to capitalize on the wide range / detailed information which a survey collects and at the same time, the wide coverage a census data set affords. In this context, census data are referred to as *auxiliary data.* In addition to census data, data from administrative reporting systems provide auxiliary data for SAE.

There is a wide range of small area estimation techniques, and the choice usually depends on what is being estimated, what types of auxiliary data sources are available and the desired level of disaggregation. Three general approaches are direct survey estimation, SAE using auxiliary information and SAE using regression-based models. Commonly used methods[28] include broad area ratio estimation, calibration methods, weight reallocation, empirical best linear unbiased prediction (EBLUP), and other regression-based methods. Figure 3.2. provides a relationship diagram of these methods.

---

[28] For detailed discussion on these techniques, refer to ADB (2020) *Introduction to Small Area Estimation Techniques: A Practical Guide for National Statistical Offices.* https://www.adb.org/publications/small-area-estimation-guide-national-statistics-offices

Figure 3.2: SAE methods
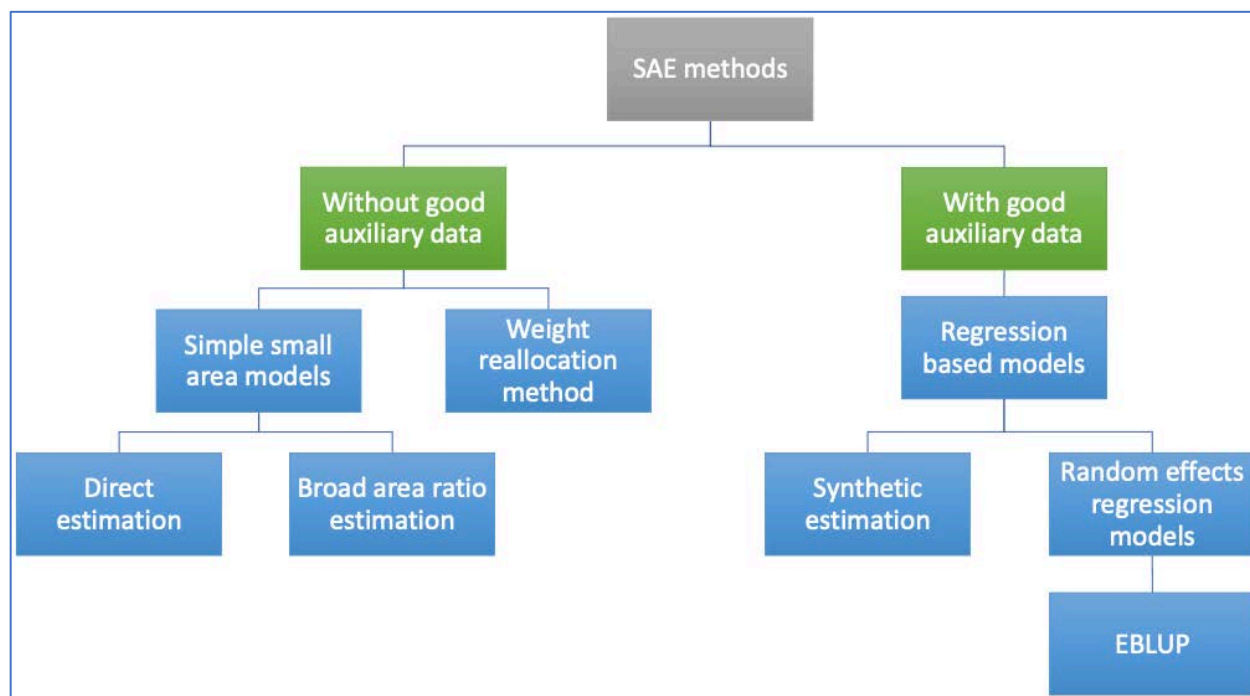
SAE techniques can be applied to further disaggregate SDG indicators that are traditionally compiled from surveys. Some illustrations on poverty, unemployment and health indicators are provided below.

Poverty-related indicators

A relatively well-developed application is the disaggregation of **poverty-related indicators**. For instance, a household income and expenditure survey (HIES) or a living standards survey (LSS) are used to compile statistics on people living below a national or international poverty line. Like many household surveys, a typical HIES or LSS have sample sizes large enough to provide reliable estimates when poverty statistics are summarized by population groups (e.g., gender, age, occupation of an individual or household head) at the national or regional level but not necessarily reliable when disaggregated at finer levels. One of the commonly used SAE technique for poverty estimation is the *Empirical Best Linear Unbiased Estimator* (EBLUP).

Box 3.3: How to estimate EBLUP?

First, one calculates poverty statistics for the desired set of small areas, using data from surveys. Second, one finds potential covariates, available at the small area level, for the calculated poverty statistics from auxiliary data. Third, the survey-based poverty statistics compiled at the small area level is regressed on the potential covariates. Lastly, the weighted average of the poverty statistics compiled in the first step and predicted poverty statistics from the third step is taken, where the weights are inversely proportional to the variance of the model error and sampling error, respectively. Such approach is feasible when the following criteria are met: (i) all small areas are

---

represented in the survey; and (ii) there are auxiliary data available at the desired small area level, and at the same time correlated with the poverty indicator being estimated.

If any of the criteria for EBLUP is not met, an alternative is the *poverty mapping method* popularized by the World Bank, which is also referred as the Elbers, Lanjouw, and Lanjouw (ELL) method. The method entails integrating income or expenditure data from HIES or LSS with auxiliary data from population censuses (ideally collected in the same year). It employs the survey data to build an econometric model by regressing the household-level data on income or expenditure collected from surveys with income correlates that are both available from surveys and census. The parameter coefficients derived from the econometric model are then applied to the census values of the income correlates. This method yields income or expenditure predictions for every unit in the census data.

Box 3.4: Poverty Maps of Uganda (2018)

The Uganda Bureau of statistics (UBOS), United Nations International Children's Emergency Fund (UNICEF), and the World Bank created new poverty maps at the sub-county level based on the ELL method. The maps include *child poverty estimates* across all geographic regions. The SAE estimates were based on the 2012-13 Uganda National Household survey and the 2014 National Population and Housing Census.

The methodology, mapping exercise and validation and results are presented @
http://documents1.worldbank.org/curated/en/456801530034180435/pdf/Poverty-Maps-Report.pdf

Poverty mapping exercise in Palestine (2019)

A study of SAE methods to estimate *poverty rates* and *gaps* in localities (small area) in Palestine used the 2016/17 Palestinian Expenditure Consumption Survey and the Population Census of 2017. The study concludes that "the obtained Census Best/Bayes Predictor estimates based on the nested error model take reasonable values for all the localities of interest, seem to be absent of serious systematic design bias and have smaller estimated MSEs than the corresponding direct estimators for nearly all the localities, with an average MSE reduction of 84% for poverty rates and even greater reduction for poverty gaps.

https://www.unescwa.org/sites/www.unescwa.org/files/u1461/short_guide_for_small_area_estimation_in_household_surveys-poverty_mapping_in_palestine.pdf

---

Unemployment rates

Box 3.5: SAE applied to unemployment rates in Indonesia (2015)

The Sakernas or National Labour Force Survey of Indonesia is conducted based on a quarterly rotating panel survey. Because groups differ according to their time-in-panel and observation strategy, it leads to bias. In addition, sample size is not adequate to obtain reliable direct estimates of indicators at the district level (small area). Currently, a small area estimation (SAE) model that accommodates the bias component due to the rotation is applied but only assumes the effect over time follows a random walk process, so it is necessary to develop a model that is more general. Estimates from an SAE model for rotation group level, combined with the time-series multi-level model and the Rao-Yu model have produced data to estimate a quarterly unemployment rate at the district level.

<u>*Health-related indicators*</u>

Small area estimation is also being used to compile more disaggregated data on **health indicators** compiled through surveys. For example, in addition to providing disaggregated data on income- or expenditure-based indicators like poverty and inequality, the ELL method has been used to produce disaggregated data on non-income indicators like stunting and wasting, as well as health and disease maps, in a number of countries (Van der Weide, 2017). In the United States, small area estimation technique is employed to compile disaggregated data on the prevalence of childhood obesity (Li, Kelsey and Zhang, 2009; Zhang, Onufrak, Holt and Croft, 2013). Other notable SAE applications include compiling disaggregated data on the number of underweight children aged 0–5 years in municipalities and cities in two different regions (Relente 2010 and Aracid, 2014), the proportion of underweight children aged 0–5 years (Arlan, 2016), and the proportion of Vitamin A-deficient children aged 6 months–5 years (Abitona, 2011), and prevalence of maternal mortality (Nuestro, 2014; Fernandez, 2014).

Potentials

Overall, small area estimation could be a potentially more cost-effective analytical tool for providing disaggregated estimates for a wide range of SDG indicators which are traditionally compiled from surveys alone. However, SAE requires availability of auxiliary data and several techniques employ normative assumptions, validity of which need to be carefully examined.

Thus, before one embarks on a small area estimation exercise, it is important to have clarity on what type of disaggregated data are needed, and why such type of data is needed; what are the key policies, funding decisions, or questions that disaggregated statistics can address. SAE should be done when there is a valid clamor for disaggregated statistics but no other proxy data at the small area level are available for use in policymaking and program planning. In some instances, volumes of data are readily available, yet remain untapped. In relation to the SDG data requirements for disaggregated indicators, these information and conditions are articulated in the global commitments, but some work still needs to be done at the national level.

Limitations

It is not always the case that implementing small area estimation provides reliable estimates at all desired levels of disaggregation. In some instances, data producers publish estimates calculated directly from surveys at the small area-level, but with notes or caveats stating that caution should be taken when examining small areas for which reliability of survey estimates fall below a pre-specified threshold. On the other hand, the application of SAE becomes more compelling when many units of the desired level of disaggregation have direct survey estimates falling below a pre-specified reliability threshold.

National statistical activities require considerable financial and technical resources and embarking on SAE studies and implementation would need resources even if it is regarded as a more cost-

efficient tool than increasing the sample size of surveys. Because different stakeholders may need varying levels of disaggregation, it is important to identify whose data needs can be addressed most efficiently by conducting SAE through consultations. As SAE is not a simple solution that can address all disaggregated data requirements of every user, consultation is a good opportunity to communicate the limitations of SAE to target users.

In the context of SDG monitoring, it is important to note that SAE is not a silver bullet that can address all of the disaggregated data requirements. SAE techniques start with sample survey data would thus not be relevant for indicators which are not derived from surveys. There are also instances when fundamental issues such as absence of consistent conceptual definitions, etc., remain unresolved, which prevents the compilation of disaggregated data.

Recommended Reading
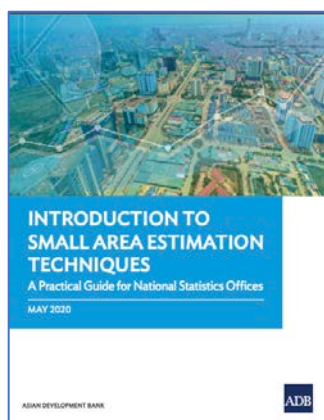
Asian Development Bank (2020. Introduction to small area estimation techniques: A practical guide for national statistics offices.
https://www.adb.org/sites/default/files/publication/609476/small-area-estimation-guide-nsos.pdf
This guide provides an introduction to basic SAE techniques and describes the implementation of these techniques using R software.

First, it provides a rationale for conducting SAE in the context of meeting the disaggregated data requirements of the SDGs. Second, it walks readers through the basics of data management and analysis using R. Third, it presents various SAE techniques, from simple calibration methods to progressively more sophisticated modeling procedures. The discussion of each SAE technique includes several examples to demonstrate step-by-step implementation of the underlying procedure. Themes of calculation examples are drawn mainly from commonly collected household surveys (e.g., the Family Income and Expenditure Survey, the Labor Force Survey, and the Demographic and Health Survey). Software implementation using R is provided throughout the discussion.

A Toolkit on Small Area Estimation for the SDGs
https://unstats.un.org/iswghs/task-forces/task-forces-round2/

The Joint Intersecretariat Working Group on Household Surveys and Interagency Expert Group on Sustainable Development Goals (IAEG-SDGs) Task Force on Small Area Estimates led by the UNSD and members coming from NSOs, international agencies, academia, and NGOs are developing a toolkit on SAE for the SDGs. The toolkit includes an overview of small area estimation methods, step-by-step guides on using small area estimation methods, country examples and case studies, visualization and communication of small area estimates, and software packages for SAE. The tentative title of the toolkit is SAE4SDG.

…

3.5 Sources of disaggregated data: "Big Data", Geo-Spatial Data and related others

Digital footprints, referred to as "big data", offer possibilities to address SDG data disaggregation gaps or to complement statistics generated from traditional data sources.

These sources can be classified as shown in Table 3.5.1.[29]

Table 3.5.1: A taxonomy of big data

Type	Sub-types	Examples
Social Networks (human-sourced information)	Social networks	Facebook, Twitter, Tumblr
	Blogs and comments	
	Pictures	Instagram, Flickr, Picasa
	Videos	YouTube, Tik-Tok
	Internet searches	
	User-generated maps	
Traditional business systems (process-mediated data) Some sources may fall into the category of "administrative data".	Data produced by public agencies	Medical records
	Data produced by businesses	Commercial transactions: scanner data
		Banking/stock records
		e-Commerce
		Credit cards
Internet of things (machine generated data)	Data from sensors	**Fixed sensors**: home automation, weather/pollution sensors, traffic sensors/webcam, security/surveillance videos
		Mobile sensors (tracking): mobile phone location, cars, satellite images
	Data from computer systems	Logs, web logs

"Big data" uses for SDGs monitoring: Examples

Generally, the use of big data for producing official statistics is still limited but increasingly being explored[30]. Applications to the production of SDGs monitoring indicators drive some of the methodological work in this area. Some examples of ongoing work that facilitate the

---

[29] Source: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2019/ECE_CES_2019_41.pdf
[30] Refer to: E/CN.3/2020/24 Report of the GWG on Big Data for Official Statistics. https://unstats.un.org/unsd/statcom/51st-session/documents/UN_BigData_report_v6.0-E.html

disaggregation of SDG are highlighted in Box 3.6. Hyperlinks to available details are provided but not discussed in the text of this Guidebook.

Box 3.6: Examples of use of big data for producing SDG indicators

Statistics Korea and SKT have launched a [data service for population movement between regions](), generating statistics on traffic volume by gender/age

BPS Statistics Indonesia and a wide range of partners are using [mobile positioning data]() to increase coverage and granularity for tourism statistics (SDG target 8.9 and 12.b), for commuting and internal migration statistics (SDG targets 10.7 and 11.a)

[Ideamaps Network]() is integrating four current approaches to *slum area mapping* and Flowminder Foundation demonstrated use of [spatial models for slum area mapping]()

The Philippines is [refining methodology]() that relates the *rural access index* (SDG indicator 9.1.1) based on GIS mapping and software to survey-based poverty estimates.

[World Data Lab]() data models track the SDGs in real time at disaggregated levels, such as through a Poverty Clock that disaggregates poverty headcount by gender, age and rural/urban. Various examples are presented in [https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%207.b_World%20Data%20Lab.pdf]()

The examples provided above are illustrative of the increasing initiatives in the use of big data for official statistics, in general, and for producing disaggregated data needed for SDG monitoring. A methodological challenge is that of providing a "proof-of-concept" for scalability of some of the results. An example of how one has been developed in relation to producing disaggregated poverty estimates published by ADB (2020)[31] is briefly described in Box 3.7 below.

Box 3.7: Studies on scalability: Producing more disaggregated poverty estimates

Under its 'Data for Development' knowledge and support technical assistance project, the Asian Development Bank (ADB) is keeping tabs on relevant initiatives using big data so that countries and their stakeholders can have a more nuanced understanding on the scalability of such initiatives. For example, pilot work in Thailand and the Philippines has meshed satellite imagery data, with data from censuses and surveys to improve the accuracy and level of disaggregation of small area estimates of poverty incidence. Prior to these, Thailand and the Philippines combined data from sample surveys with census data through SAE models to yield more disaggregated poverty estimates.

As discussed in the section on SAE, the detailed survey data to construct a "model" of expenditure, income (or any other household- or individual-level indicators of welfare) as a function of variables that are common to both the household survey and the census. Estimates from the small area estimation model, along with census data on the variables that help predict welfare, are used to get predicted values of poverty indicators for every household in the country. These predicted values are then used to estimate poverty incidence and other poverty measures

---

at a much more granular level. The ADB project explored the use of luminosity and other satellite imagery data to complement the use of survey and census data for small area estimation.

A step-by-step approach for the use of big data in modelling for official statistics is presented in Box 3.8.

Box 3.8: Enhanced step-by-step approach for the use of big data in official statistics

These steps outline a framework when making a decision with regard to when, how and for what using big data in official statistics, based on the study of Buono, et al. (2018).

Step 1: specify the statistical problem (e.g., nowcasting)

Step 2: define an appropriate IT environment for collecting, storing, treating and analyzing big data

Step 3: consider Big Data potential usefulness within the statistical exercises as defined in step 1

Step 4: discuss issues related to the design of the exercise: possible big data sources, definition of the type of variables/information to be extracted from big data, etc.

Step 5: assess big-data accessibility and the quality of envisaged big-data sources

Step 6: explore big data specific features and on possible ways to move from unstructured to structured data

Step 7: consider approaches to replace missing observations as well as to remove and correct outliers

Step 8: consider ways for filtering out seasonal and other very short-term components from big data

Step 9: determine presence of bias in big data and the ways to correct it

Step 10: analyze the design of the econometric modelling strategy in case of the availability of numerical big data structured in time series form

Step 11: consider the statistical exploratory techniques to be used in case of nonnumerical or numerical but not structured big data

Step 12: evaluate the results mainly obtained in steps 10 and 11 (choice of benchmark model or nowcasts, evaluation metrics, criteria for assessing the significance of the accuracy and/or timeliness gains, the statistical and economic validity of the results, etc.)

Step 13: consider the overall implementation of big data-based results within the regular production and dissemination process

Source: Buono, D., G. Kaptanios, et al. Paper prepared for the 16th conference of IAOS, September 2018. https://www.oecd.org/iaos2018/programme/IAOS-OECD2018_Buono-Kapetanios-Papailias-Marcellino-Mazzi.pdf.

"Big data" uses for SDGs monitoring: Potentials and (current) limitations

Findings from a 2017 survey of NSOs in Asia and the Pacific presented in the ADB brief entitled, *Readiness of National Statistical Systems in Asia and the Pacific for Leveraging Big Data to Monitor the SDGs*[32] provide insights from responding NSOs on factors affecting its use. These include *access* to big data sources, *technological infrastructure* requirements, acquiring new *skills*, and concerns relating to *data privacy* and *security*.

On the other hand, the 2019 survey conducted by the United Nations Global Working Group on Big Data for Official Statistics presented in the report *Global Assessment of Institutional Readiness for the use of Big Data in Official Statistics*[33] show that a large proportion of the 100 responding NSOs are already embracing big data but may require further information, guidance, development and knowledge to remove barriers to working with big data. Table 3.5.2 summarizes the current status of NSOs vis-a-vis the factors and barriers in utilizing big data. The analysis presented in the report provide details on what are covered under the four key points.

Table 3.5.2: Key factors and barriers in utilizing big data for official statistics, including SDGs indicators

| Key Points	Status	Barriers
1- Strategic coordination	Strategic coordination capacities are fairly established. Many NSOs are actively engaged in big data projects. *Ethics and quality frameworks* are fairly established. Most NSOs view *coordination with Big Data source owners* inside their NSS as the lowest challenge.	Only a third of all NSOs have *overarching big data strategies* in place and *Chief Data Officers* only exist in some NSOs. The biggest challenge for NSOs is the *collaboration with Big Data source owners outside the government*.
2- Legal framework	Overall, respondents are aware of the fundamental role legal frameworks play in establishing big data projects. Many NSOs appear to have well developed legal frameworks that penalize *data disclosures* and allow *accredited access* to their data.	Legal frameworks are still insufficient to *regulate big data applications*. Only a small share of NSOs rely on legal frameworks that *guarantee access to big data*.
3- IT infrastructure	The analysis shows a heterogeneous picture over the IT infrastructure. NSOs stated that basic IT infrastructure such as power supply, and air-conditioning	IT infrastructure appears as central barrier to develop big data capacity; *onsite and offsite storage capacity* needs improvement for many. Only a few NSOs

---

[32] Refer to: https://www.adb.org/publications/national-statistical-systems-big-data-sdgs. The report lists ongoing/completed big-data related initiatives in Asia and the Pacific as of 2017.
[33] Refer to: https://unstats.un.org/unsd/statcom/51st-session/documents/UN_BigData_report_v6.0-E.html

Key Points	Status	Barriers
	mostly met their needs, but they outlined struggles with storage facilities and computing power.	consider *cloud storage* a relevant option.
4- Human resources	NSOs recruit significantly *more analysts than data scientists* and prioritize *up-skilling over hiring external staff* to perform big data/data science techniques.	Most NSOs lack a *competency framework* to develop new skills to cope with big data (mobile phone, geospatial data) and new methodologies (machine learning).

Resources

Recommended Reading
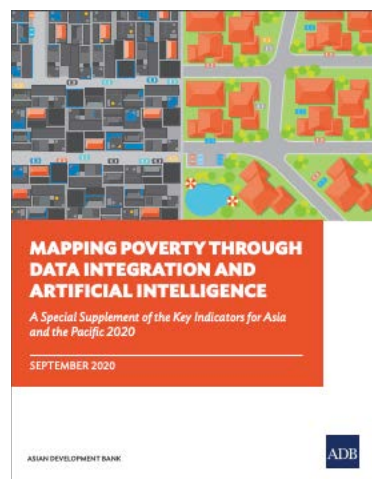
Asian Development Bank (2020). Mapping poverty through data integration and artificial intelligence: A special supplement of the Key Indicators for Asia and the Pacific 2020 https://www.adb.org/sites/default/files/publication/630406/mapping-poverty-ki2020-supplement.pdf
This publication presents initial results of studies under the ADB knowledge initiative Data for Development, which aims to strengthen the capacity of NSOs in the Asia and Pacific region to meet the increasing data demands for effective policymaking and for monitoring development goals and targets.

The approach demonstrates use of high-resolution satellite imagery, geospatial data, and powerful machine-learning algorithms to complement traditional data sources and conventional survey methods to estimate the magnitude of poverty. Statisticians from ADB's Statistics and Data Innovation Unit within the Economic Research and Regional Cooperation Department worked with the Philippine Statistics Authority, the National Statistical Office of Thailand, and the World Data Lab to examine the feasibility of poverty mapping using satellite imagery and associated geospatial data.

Marchetti, *et al.* 2015. Small Area Model-Based Estimators Using Big Data Sources. Journal of Official Statistics, Vol 31, No 2 @
https://lucapappalardo.files.wordpress.com/2014/08/jos2015.pdf

This article discusses that integrating statistical research in small area estimation with research in social data mining built on Big Data, can present innovative ways of addressing the challenge of generating timely, accurate monitoring of social indicators, such as poverty or inequality, on a granular spatial and temporal scale in official statistics. The article recommends three approaches to use Big Data along with small area estimation methodology and explains how Big Data can illustrate facets of well-being and other socioeconomic trends.

UNSD. 2016. Report of the 2015 Big Data Survey
https://unstats.un.org/bigdata/documents/reports/GWG%20Background%20document%20-%202016%20-6-Report-of-the-2015-Big-Data-Survey-E.pdf

This report provides the outcome of a survey conducted from June to August 2015 among NSOs of 93 countries. The survey aims to examine the situation regarding the steps undertaken by statistical agencies in relation to Big Data. It also asks the strategic vision of NSOs and their practical experience with Big Data. Participating countries reported the need for training and capacity-building on Big Data topics (i.e., methodologies, estimation, and a quality framework) and conducting more pilot projects especially for developing countries. Further, NSOs relayed that easier access to Big Data can lower the threshold for undertaking a Big Data project.

United Nations. E/CN.3/2020/24. Global assessment of institutional readiness for the use of big data in official statistics @ https://unstats.un.org/unsd/statcom/51st-session/documents/UN_BigData_report_v6.0-E.html

This document reports the outcomes of one of the projects of the UN GWG Task Team on Training, Competencies and Capacity Development, i.e., An assessment of NSO readiness for the use of big data in official statistics. The findings demonstrate that a large proportion of NSOs that participated in the survey are already adopting big data or data science. In general, the findings present a positive scenario in terms of confirming that the required foundations are in place and demonstrates the determination among NSOs to integrate big data into their system.

3.6 Sources of disaggregated data: Data integration

Data integration is the process of combining or linking data from different sources to produce more comprehensive and disaggregated statistics and information. Most often, the data sources (i.e. census, surveys, administrative data, geospatial information, etc.) are designed and collected primarily without the intention of being used together. The benefits of bringing together information from different sources include the ability to answer a broader range of questions, examine underlying relationships between cross-sections of society and produce more timely, disaggregated statistics and at a greater frequency than traditional approaches alone. In addition, it is considered less time consuming and less costly when compared to other methods such as surveys and can also reduce respondent burden. However, national statistical offices also face challenges when integrating data such as access to data, interoperability, technical capacity and significant time and resource commitments.

There are many possible types of data integration. Five common types are: administrative sources with survey and other traditional data; new data sources (such as big data) with traditional data sources; geospatial data with statistical information; micro level data with data at the macro level; and validating data from official sources with data from other sources. Examples are provided in the Box 3.9 below.

Box 3.9: Some examples of data integration outputs

- o   an integrated dataset that serves as an input to produce official statistics

- o   a statistical model developed and produced using different sources to produce model-based information

- o   a dataset integrated for the purposes of micro-validation when some rules are defined to check the validity of the data in one dataset compared to another one

- o   missing values imputed in a dataset using another dataset as the source for imputation

- o   datasets combined to produce a sampling frame for a survey

- o   data from several subject-matter domains combined into one dataset that is the basis for the production of statistics (example: national accounts)

- o   datasets from different subject-matter domains compared to check the quality and the validity of information produced (macro-validation)

- o   input from several sources integrated into one dataset to provide microdata files for the researchers for scientific purposes

- o   different sources used to apply proper statistical disclosure control methods on a microdata set.
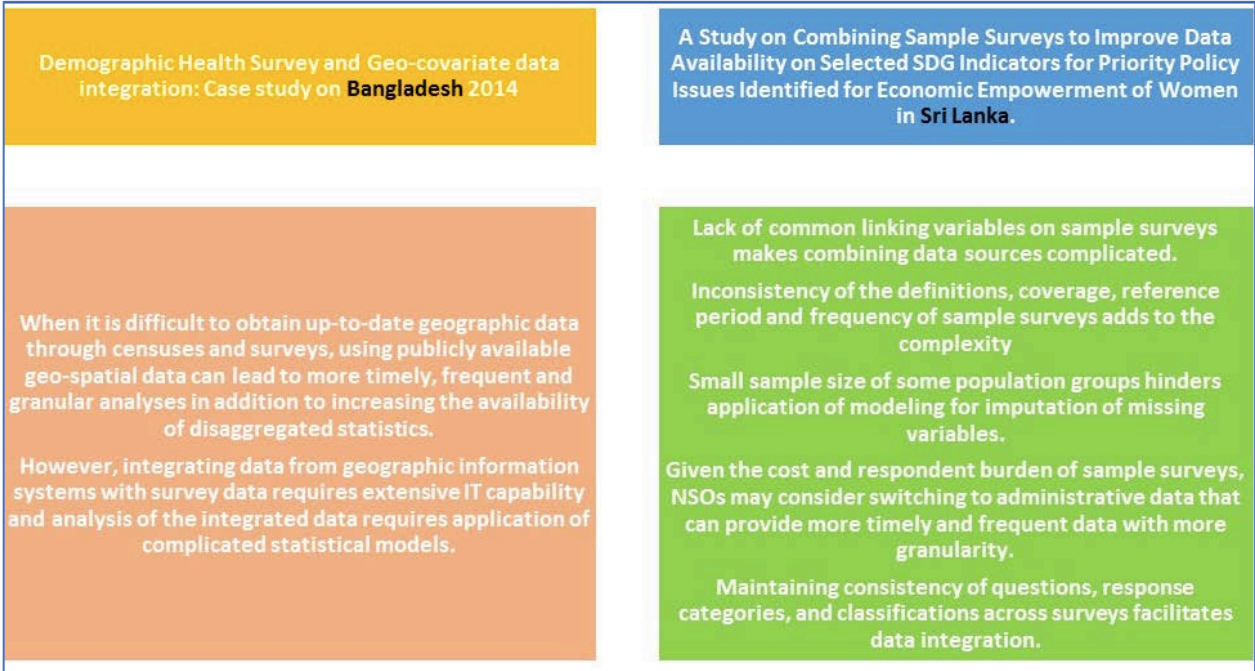
Example case studies

In an effort to build technical capacity, UN ESCAP supported two case studies on data integration. The first case study, from Bangladesh[34], focused on inequality and poverty indicators using non-traditional data sources (geospatial data) and a health survey. The second case study, from Sri Lanka[35], focused on women's empowerment indicators using more traditional data sources- a labour force survey and a household and income and expenditure survey.

---

[34] See: https://communities.unescap.org/system/files/report_dhs_and_geo-covariates_data_integration_bangladesh_survey_2014.pdf
[35] See: https://communities.unescap.org/system/files/improving_data_availability_for_economic_empowerment_of_women_in_sri_lanka.pdf

Figure 3.3: Bangladesh and Sri Lanka Case Studies

Demographic Health Survey and Geo-covariate data integration: Case study on **Bangladesh** 2014	A Study on Combining Sample Surveys to Improve Data Availability on Selected SDG Indicators for Priority Policy Issues Identified for Economic Empowerment of Women in **Sri Lanka**.
When it is difficult to obtain up-to-date geographic data through censuses and surveys, using publicly available geo-spatial data can lead to more timely, frequent and granular analyses in addition to increasing the availability of disaggregated statistics. However, integrating data from geographic information systems with survey data requires extensive IT capability and analysis of the integrated data requires application of complicated statistical models.	Lack of common linking variables on sample surveys makes combining data sources complicated. Inconsistency of the definitions, coverage, reference period and frequency of sample surveys adds to the complexity Small sample size of some population groups hinders application of modeling for imputation of missing variables. Given the cost and respondent burden of sample surveys, NSOs may consider switching to administrative data that can provide more timely and frequent data with more granularity. Maintaining consistency of questions, response categories, and classifications across surveys facilitates data integration.

Source: Using Data Integration to Meet the Ambitions of the 2030 Agenda blog (UN ESCAP)[36]

Resources providing guidance on applications of data integration for generating data for specific vulnerable populations are briefly described in the Box below.

---

[36] See: https://www.unescap.org/blog/using-data-integration-meet-ambitions-2030-agenda

Box 3.10:. Methods for data integration with specific applications

a-Persons with disabilities

The Model Disability Survey (MDS), developed by WHO and the World Bank, is a general population household survey that provides detailed and nuanced information about how people with and without disabilities conduct their lives and the difficulties they encounter, regardless of any underlying health condition or impairment. The brief version of the MDS can also be integrated in other surveys. For example, it was implemented as a module with the Gallup World Poll in India, Tajikistan and Laos. The Gallup World Poll is the only global study of its kind and contains dozens of questions on a variety of environmental factors relevant to the disability experience like economic empowerment, access to clean water, access to nutritious food, citizen engagement, and public safety. The implementation of the Brief MDS with the Gallup World Poll was implemented in these three countries as a means of testing a new efficient way of obtaining high-quality, standardized and comprehensive disability data globally. The combination of the Brief MDS with the World Poll variables not only allows for the disaggregation of key World Poll indicators by level of disability, but it also allows to go beyond disaggregation and analyze the complex ways in which all of these societal factors influence disability, and vice versa.

b-Migration

UNECE produced the *Guidance on Data Integration for Measuring Migration*[37], outlining ways that data integration is used to produce migration statistics with thirteen case studies on data integration in various national contexts and best practice principles for national statistics offices.

c-Refugees/IDPs

UNHCR's[38] International Recommendations on Refugee Statistics (IRRS) and the International Recommendations on Refugee Statistics (IRIS) provide general recommendations on data integration, while the Compilers Manual presents the main steps related to data integration and an overview on the possible data integration techniques which can help and guide practitioners in NSOs interested in applying data integration to refugee and IDPs statistics.

d-Urban/slum areas

UN-Habitat uses the DHS, MICS and LSMS modules for household characteristics with regard to measurement of some SDG indicators[39].

General Guidance on Data Integration Methods

Several guidance documents and reviews of data integration have been developed by national statistics offices and other regional organizations.

---

[37] *Guidance on Data Integration for Measuring Migration* (UNECE): http://www.unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20186.pdf
[38] Additional information: https://www.unhcr.org/refugee-statistics/methodology/
[39] See: http://unhabitat.org.ir/wp-content/uploads/2019/03/SDG-11.1.1-Newsletter_2.1.pdf

Statistics New Zealand. *Data integration manual:2nd edition*[40], which includes best practices and insights gained from their over 10 years' experience of integrating data. The document outlines basic concepts, theory and processes involved in data integration along with practical advice and operational aspects, including privacy concerns.

UNECE and The Conference of European Statistics conducted an in-depth review of data integration[41] aimed to address a general lack of guidance and overview of experiences. The document outlines types of data integration, a general framework and broad approach as well as opportunities, challenges, skills and resources needed. It includes several types of data integration, such as survey and administrative sources, survey and new data sources (including Big Data), traditional sources with geospatial information, and integrating data for validating official statistics.

UNECE. *A Guide to Data Integration for Official Statistics (version 2.0)[42],* providing practical advice and information to advance data integration activities by statistical organisations and information about issues that statistical organisations have or should consider in work on data integration.

UN ESCAP. *Integrated Statistics: A journey worthwhile*[43], focusing in a quick reference document on process integration and data integration, addressing delivery issues relevant to each, while also outlining different sources common for official statistics production.

FAO. Details techniques and guidelines for using remote sensing and other agricultural censuses and surveys for agricultural statistics including the Handbook on remote sensing for agricultural statistics, the Technical Report on Reconciling Data from Agricultural Censuses and Surveys and additional publications on remote sensing and geographic information and data integration using small area estimations[44].

The Global Statistical Geospatial Framework[45], developed by the UN Global Geospatial Information Management (UN-GGIM) Secretariat, facilitates the integration of statistical and geospatial information through the application of five Principles and supporting key elements. These data can also be integrated with statistical, geospatial, and other information to inform and facilitate data-driven and evidence-based decision making and in particular for the 2030 Agenda and the SDGs.

The Territorial Dimension in SDG Indicators: Geospatial Data Analysis and its Integration with Statistical Data[46] (UN-GGIM Europe) report addresses the territorial dimension of the SDGs and the contribution of geospatial information integrated with statistical data for disaggregated, aggregated, mapped and visualized information.

---

[40] *Data integration manual: 2nd edition* (Statistics New Zealand): http://archive.stats.govt.nz/methods/data-integration/data-integration-manual-2edn.aspx#gsc.tab=0. Additional information: https://www.stats.govt.nz/integrated-data/
[41] See: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/bur/2017/February/02_in-depth_review_data_integration_final.pdf. Additional information: https://www.unece.org/stats/ces/in-depth-reviews/geospatial2.html
[42] See: https://statswiki.unece.org/spaces/flyingpdf/pdfpageexport.action?pageId=129171769
[43] See: https://www.unescap.org/sites/default/files/Stats_Brief_Issue19_Jul2019_Integrated_Statistics.pdf
[44] See: http://gsars.org/en/tag/geoinfo/
[45] See: http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf
[46] See: https://un-ggim-europe.org/wp-content/uploads/2019/05/UN_GGIM_08_05_2019-The-territorial-dimension-in-SDG-indicators-Final.pdf

What's Next

The next chapter highlights the need to assess data reliability of disaggregated information drawn from survey microdata, inform the users of the constraints of estimates or resort to withholding the publication of disaggregated data when its reliability does not meet a pre-determined threshold (Section 4.1.2). The chapter also discusses that disaggregated data sourced from specific household surveys, i.e., DHS, MICS and RHS conducted in more than 100 countries, serve as inputs in the Health Equity Assessment Toolkit (HEAT) of the World Health Organization (WHO), which examines within-country health inequalities (Section 4.4.1).

Chapter 5 provides a specific case study in the Philippines on how disaggregated data on poverty collected from various data sources led to actual policy uses of disaggregated statistics. The data sources include the Family Income and Expenditure Survey and Census of Population and Housing of the Philippine Statistics Authority, the Community Based Monitoring System, and the National Household Targeting System for Poverty Reduction (NHTS-PR) of the Philips partment of Social Welfare and Development (Section 5.3.1).

Annex- Online resources from Workshops on SDG Disaggregation

International Workshop on Data Disaggregation for SDGs, 28-30 January 2019, Bangkok
https://unstats.un.org/sdgs/meetings/sdg-inter-workshop-jan-2019/

Topic/Area	Online Resource
Overview of Data Disaggregation for the SDGs, assessments, general approaches	UNSD: Data disaggregation and the Global Indicator Framework. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%202.a_UNSD%20IAEG.pdf
	Cambodia experience in producing disaggregated data for SDGs. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%202.b.1_Cambodia_Experience%20in%20producing%20disaggregated%20data%20for%20SDGs.pdf
	Data disaggregation practice on national SDG implementation in Ethiopia. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%202.b.2_%20Ethiopia%20Data%20Disaggregation%20Practice%20on%20National%20SDG.pdf
	Data disaggregation for SDGs: Turkmenistan's experience. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%202.b.3_Turkmenistan_Bang_270119%20ENG%20PRINT.pdf
	Data disaggregation for SDGs- United Kingdom. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%202.b.4_UK_DataDisagregation_ONS_DFID.pdf
Integrating policy demands on inclusive development with data	OHCHR: Human rights-based approach to data. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%203.a_OHCHR%20HRBAD%20data%20disaggregation.FINAL.pdf

Topic/Area	Online Resource
	EPIC: A tool for policy-data integration https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%203.b_ESCAP_Data%20Disaggregation_Jan2019.pdf
	EPIC Case Study- Samoa: Education sector. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%203.c.1_Samoa_Presentation_5.pdf
	EPIC Case Study- Philippines: Women's Economic Empowerment. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%203.c.2_%20Philippines%20Presentation_EPIC.pdf
Sources: Censuses and Household Surveys	UNSD: Using census for producing disaggregated data. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%206.a_CensusDisaggr-UNSD.pdf
	UNICEF: Producing disaggregated data from household surveys (highlighting MICS). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%206.b_UNICEF_DA%20Data%20Disaggregation%20Bangkok%20Jan%202019.pdf
	Mexico: Household surveys for producing disaggregated data. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%206.c.1_Household%20survey%20data_MexicanCase_29Jan2019.pdf
	Ghana: Ghana's experience (utilizing various data sources). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%206.c.2_Ghana_DataDisaggregation_Thailand_Gh_28012019.pdf
Sources: Administrative data	UNSD: Use of administrative sources for producing disaggregated data (highlighting challenges). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%205.a_UNSD.pdf

Topic/Area	Online Resource
	Republic of Korea: examples of general initiatives and challenges. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%205.b.1_Korea%20Using%20administrative%20data_KOSTAT.pdf
	Rwanda: Using administrative data sources for producing disaggregated data (highlighting administrative systems. Potential uses and challenges). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%205.b.2_Rwanda_presentation_administrative_data_for_sdgs.pdf
Sources: SAE	ADB: Small area estimation and big data (theory and illustrations of SAE). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%207.a_ADB_Small%20Area%20Estimation%20-%20jan29.pdf
	Thailand: Illustrating poverty mapping. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%207.c.1_Country%20Presentation%20Thailand.pdf
	Philippines: Efforts in monitoring SDG with disaggregation (including overview of SDG implementation framework with in the NDP and highlighting existing efforts on SAE). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%207.c.3_SAE%20Presentation%20Philippines_29January2019%20rev.pdf
	Indonesia: Small area estimation study for providing and disaggregating SDGs indicators information on various studies on using SAE for producing six (6) SDG-related indicators). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%207.c.2_Indonesia-SAE%20for%20Data%20Dissagregassion-January2019.pdf
Sources: Big data	ADB: Small area estimation and big data (illustrations of use of big data in SAE). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%207.a_ADB_Small%20Area%20Estimation%20-%20jan29.pdf

Topic/Area	Online Resource
	Case study- Uganda. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%2010.b.1_Inequality_In_uganda.pdf
Inequality and LNOB analysis: Gender equality and multiple deprivation assessment	UN Women: Gender data and multilevel disaggregation: an LNOB perspective to SDG monitoring (overview of gender-relevant and gender-specific SDG indicators, disaggregation dimensions, illustrations). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%208.a_UNWomen_Gender%20data%20and%20multi%20level%20disaggregation.pdf
	Uganda: User engagement to inform the production of gender-related SDG data (national priority gender and equality indicators). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%208.b.1_Uganda-User%20Engagement-Case%20for%20NPGEIs-Bangkok%20Jan%202019.pdf
	Bangladesh: Integrated approach to producing and using gender statistics for SDG monitoring. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%208.b.2_Bangladesh_Country%20Presentation%20Thailand.pdf
	Viet Nam: SDG from perspective on gender equality (highlighting development of gender indicator framework). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%208.b.3_Vietnam_SDG%20from%20gender%20equality%20perspective%20in%20Vietnam.pdf
Inequality and LNOB analysis: Health equity assessment	WHO: Work on data disaggregation (with introduction to HEAT, HEAT plus, other resources and illustrations). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%2010.a_WHO%20data%20disaggregtion%20work-Jan%202019%20-%20IAEG_SDG_Bangkok.pdf
	Case study- Uganda. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%2010.b.1_Inequality_In_uganda.pdf

Topic/Area	Online Resource
	Case study- Indonesia. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%2010.b.2_indonesia%20experience-AH_nk_rev.pdf
Inequality and LNOB analysis: Education equity assessment	UNESCO Institute of Statistics: Assessment of equity in education (highlights sources, guidelines on equity assessment and illustrations). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%2011.a_UNESCO_ESCAP%20data%20disagreegation.pdf
	Rwanda: Education equity measurement. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%2011.b.1_Rwanda_SDG%20data%20disaggregation_Education.pdf
	Nepal: Education equity assessment- practices and experience. https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%2011.b.2_Nepal_Education_Equity.pdf
	Philippines: Education equality in the Philippines (features analyses of data from household surveys and administrative systems). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%2011.b.3_Philippines___Education%20Equality%20AssessmentFINAL4.pdf
Enabling environment and capacity development	Uganda: User engagement to inform the production of gender-related SDG data (highlighting the process for adoption of a national priority gender and equality indicators). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%208.b.1_Uganda-User%20Engagement-Case%20for%20NPGEIs-Bangkok%20Jan%202019.pdf
	Bangladesh: Integrated approach to producing and using gender statistics for SDG monitoring (highlighting engagement with ministry of women in strengthening administrative

Topic/Area	Online Resource
	data sources). https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%208.b.2_Bangladesh_Country%20Presentation%20Thailand.pdf

ESCWA Regional Workshop on Data Disaggregation for SDGs Indicators, 19-21 November 2019, Istanbul
@ https://www.unescwa.org/events/regional-workshop-data-disaggregation-sdgs-indicators

Topic/Area	Online Resource
Overview of Data Disaggregation for the SDGs, assessments, general approaches	UNSD: Data disaggregation and the Global Indicator Framework @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/1.1_unsd.pdf
	ESCWA: Data disaggregation for the SDGS (illustrating practices in the region) @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/1.1_escwa.pdf
	FAO: Disaggregation of the SDG indicators related to food and agriculture @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.1_fao.pdf
	Sudan: SDG data gap analysis @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/1.2_sudan.pdf
Integrating policy demands on inclusive development with data	ESCWA module on how disaggregated data shape policy formulation @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/3.1_escwa.pdf
	Introduction to EPIC (Every Policy is Connected) @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/3.1_epic.pdf
Sources: Censuses and Household Surveys	Palestine: Collecting and compiling disaggregated data from surveys and administrative records @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/1.2_palestine.pdf

Topic/Area	Online Resource
	Oman: Using the census for disaggregating SDG indicators @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/1.2_oman.pdf
	Sudan: Surveys and censuses for vulnerable groups analysis @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.2_sudan.pdf
	Morocco: Women and girls in the census @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.2_morocco.pdf
	Palestine: SDG 5 and disability @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.2_palestine.pdf
Sources: Administrative data	Palestine: Collecting and compiling disaggregated data from surveys and administrative records @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/1.2_palestine.pdf
Sources: SAE	Egypt: Poverty maps @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.1_egypt.pdf
Inequality and LNOB analysis	UNSD: SDG monitoring on inequality (SDG 10) @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.1_unsd.pdf
	ESCWA: Measures of inequality @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.1_escwa.pdf
	FAO: Disaggregation of the SDG indicators related to food and agriculture @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.1_fao.pdf
	ESCWA: Examples of data analysis and disaggregation @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.2_escwa.pdf

Topic/Area	Online Resource
Inequality and LNOB analysis: Health equity assessment	WHO; data disaggregation for monitoring health inequalities @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.1_who.pdf
Inequality and LNOB analysis	UNSD: SDG monitoring on inequality (focus on SDG 10) @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/2.1_unsd.pdf
Reporting and Communication	How to create key messages and use infographics @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/3.2_unsd.pdf
	Qatar: developing the national SDG database @ https://www.unescwa.org/sites/www.unescwa.org/files/u593/1.2_sdg_2018_-_qatar_data.xlsx
SDG Data Portals	Palestine: https://sdg-pcbs.opendata.arcgis.com/
	Qatar: https://sdg-en-psaqatar.opendata.arcgis.com/

References

Data2x. 2017. Big Data and the Well-Being of Women and Girls Applications on the Social Scientific Frontier https://www.data2x.org/wp-content/uploads/2017/03/Big-Data-and-the-Well-Being-of-Women-and-Girls.pdf (accessed on March 15, 2019)

Eagle,N., Macy, M., Claxton, R. 2010. Network Diversity and Economic Development. Science 21 May 2010: Vol. 328, Issue 5981, pp. 1029-1031 DOI: 10.1126/science.1186605 http://science.sciencemag.org/content/328/5981/1029 (accessed on March 15, 2019)

Gartner. n.d.. Big data. https://www.gartner.com/it-glossary/big-data/ (accessed on March 15, 2019)

Ginsberg, J., Mohebbi, M.H., Patel, R. S., Brammer, L., Smolinski, M. S. & Brilliant, L. 2009. Detecting influenza epidemics using search engine query data. Nature volume 457, pages 1012–1014 (19 February 2009) (https://www.nature.com/articles/nature07634 (Accessed on March 15, 2019)

Gao, L., Song, C, Gao, Z., Barabási, A-L. Bagrow, J. P., and Wang, D. 2014. Quantifying Information Flow During Emergencies. Scientific Reports volume 4, Article number: 3997 https://www.nature.com/articles/srep03997 (Accessed on March 15, 2019)

Inner City Fund (ICF) International. n.d.. Demographic and Health Survey (DHS) Programme. Rockville, MD, USA: ICF International. https://dhsprogram.com/ (accessed on March 15, 2019)

ICF International. n.d.. DHS Compiler. Rockville, MD, USA: ICF International. https://statcompiler.com/en/ (accessed on March 15, 2019)

International Household Survey Network (IHSN). n.d.. NADA survey cataloguing software. http://www.ihsn.org/projects/NADA-development (accessed on March 15, 2019)

Morales, L. G. G., and Orrell, T. Data interoperability: A practitioner's guide to joining up data in the development sector. Global Partnership for Sustainable Development Data (GPSDD). https://unstats.un.org/wiki/download/attachments/36143964/Guide%20to%20interoperability.pdf (accessed on March 15, 2019)

Open Data Watch. n.d.. Open Data to Support Sustainable Development Goals. Washington, DC, USA: Open Data Watch. https://opendatawatch.com/publications/open-data-to-support-sustainable-development-goals/ (accessed on March 15, 2019)

Smith C., Mashadi A., and Capra L. 2013. Ubiquitous sensing for mapping poverty in developing countries, Proceedings of the Third Conference on the Analysis of Mobile Phone Datasets. http://www.cities.io/wp-content/uploads/2012/12/d4d-chrissubmitted.pdf (Accessed on March 15, 2019)

United Nations ESCWA. 2020. A short guide for small area estimation in household surveys: illustration to poverty mapping in Palestine with expenditure survey and census data. https://www.unescwa.org/sites/www.unescwa.org/files/u1461/short_guide_for_small_area_estimation_in_household_surveys-poverty_mapping_in_palestine.pdf

United Nations International Children's Emergency Fund. n.d. Multiple Indicator Cluster Survey (MICS). New York City, NY, USA: UNICEF. http://mics.unicef.org/ (accessed on March 15, 2019)

United Nations International Children's Emergency Fund. n.d.. MICS Compiler. New York City, NY, USA: UNICEF. http://www.micscompiler.org/ (accessed on March 15, 2019)

United Nations Economic Commission for Europe (UNECE). 2013. What does "Big Data" mean for official statistics? https://statswiki.unece.org/download/attachments/58492100/Big+Data+HLG+Final.docx?version=1&modificationDate=1362939424184 (accessed on March 15, 2019)

United Nations Global Pulse. 2012. Big Data for Development: Challenges & Opportunities. New York City, NY, USA: UN Global Pulse. http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf (accessed on March 15, 2019)

United Nations Statistics Division. n.d.. Big Data Project Inventory. New York City, NY, USA: UNSD. https://unstats.un.org/bigdata/inventory (accessed on March 15, 2019)

United Nations Statistics Division. 2018. Summary of IAEG-SDGs Working Meeting on Data Disaggregation. https://unstats.un.org/sdgs/files/Data%20disaggregation%20working%20meeting%20Summary_final.pdf

United Nations Statistics Division. 2018. Overview of standards for data disaggregation. https://unstats.un.org/sdgs/files/Overview%20of%20Standards%20for%20Data%20Disaggregation.pdf

United Nations Statistics Division. 2020. Technical Report: Measuring SDGS Indicators Through Population and Housing Censuses and Civil Registration and Vital Statistics Data. https://www.unescwa.org/sites/www.unescwa.org/files/u1461/measuring_sdgs_through_phc_and_crvs_data.pdf

World Bank. n.d.. Automated Economic Analysis (ADEPT). Washington, DC, USA: World Bank. http://surveys.worldbank.org/adept (accessed on March 15, 2019)

World Bank. n.d.. Living Standard Measurement Study (LSMS). Washington, DC, USA: World Bank. http://surveys.worldbank.org/lsms (accessed on March 15, 2019)

Sustainable Development Goals and the 2020 Round of Censuses @ https://www.census.gov/content/dam/Census/library/working-papers/2018/demo/sdg-2020.pdf

Classification and Delineation of Urban Areas in a Census @ https://www.census.gov/content/dam/Census/library/working-papers/2020/demo/urban-stic.pdf

Counting the Hard to Count in a Census @ https://www.census.gov/content/dam/Census/library/working-papers/2019/demo/Hard-to-Count-Populations-Brief.pdf

Chapter 4. Analysis of Disaggregated Data in Aid of LNOB

Overview

Disaggregated data are important for informing policies and designing programs for vulnerable groups of the population, who can be further left behind in the development agenda. However, there are several points to consider in analyzing disaggregated data – availability of microdata, allocation of time and resources and data reliability – before these are translated into policy inputs.

This chapter provides an overview of *multi-level disaggregation analysis* of SDG indicators and illustrates its application through specific case studies on poverty and gender and intersectionality and the introduction of existing software on equity assessments. The chapter also demonstrates the importance of data visualization in telling stories on inequalities.

4.1 Basics

4.1.1 Examining equity issues with disaggregated data

As discussed in previous chapters, the 2030 Agenda for Sustainable Development highlights the importance of leaving no one behind. Several goals including no poverty (SDG1), zero hunger (SDG2), gender equality (SDG5), and reduced inequalities (SDG10) directly aim to tackle inequality, while other goals on good health and well-being (SDG3), quality education (SDG4), clean water and sanitation (SDG6), affordable and clean energy (SDG7), decent work and economic growth (SDG8) place emphasis on universality and inclusion, which build a strong agenda for examining equity issues.

Data help in discerning emerging policy concerns, informing policy choices and the design of public interventions, and monitoring policy implementation as well as evaluating impact. Disaggregated data helps to describe the conditions and needs of vulnerable groups in a society, e.g., youth, elderly, people with disabilities, indigenous people, people living with HIV, refugees, and ethnic minorities. Without disaggregated data, we do not know if policies and interventions are leaving the most vulnerable people even further behind. The availability of disaggregated data is imperative for examining inequities. Disaggregated data can assist in evaluating how specific demographic groups are performing in outcomes (e.g., education, labor, health). They can also help in considering whether policies are narrowing gaps, and in revealing patterns that can be masked by aggregate data. Disaggregated data can facilitate identifying needs so that resources can be allocated more equitably. Further, granular data can also help in monitoring equity in resources and outcomes.

4.1.2 Analysis objectives and plan

The availability of microdata provides flexibility to generate data and summary measures at different aggregation or disaggregation levels. If microdata are formatted consistently, specialized tools can be designed to automate the production of summary measures and serve routine purposes. For instance, consider country diagnostics of poverty, which require estimates of poverty incidence (and profiles of the poor) sourced from household surveys that collect

information on monetary indicators on living conditions (such as income and/or expenditure) and other welfare indicators. Often, such poverty analysis involves a considerable amount of time and resources to produce a set of template tables and graphs that describe poverty and inequality.

Drawing disaggregated information from microdata requires examination of data reliability. In some cases, microdata do not provide adequate information for every desired level of disaggregation. In surveys, for instance, the sample size for a desired disaggregation level may be too small to provide reliable estimates. In such cases, it is important that we draw the attention of users regarding the limitations of estimates. In other cases, data producers and compilers do not publish disaggregated data when its reliability does not meet a pre-determined threshold.

Box 4.1: Example of analysis tool: ADePT

The World Bank designed the [ADePT software](#) to simplify and speed up its analytical work on obtaining poverty profiles, making use of either individual- or household-level microdata. This approach helps minimize errors in poverty profiling, while introducing techniques and methods of applied economic analysis to a wide audience of poverty stakeholders. The ADePT software, available for free, generates a standard set of tables and graphs on poverty and inequality, data disaggregation on poverty statistics (for standard demographic groups), as well as regressions on monetary indicators of poverty (thus identifying determinants of poverty). In addition to poverty and inequality, ADePT has modules on labor, gender, social protection, education, health, food security, and agriculture. By automating the production of template tables that are commonly used for development research, the availability of a software such as ADePT gives more time for researchers to carefully examine disaggregated data and the messages it conveys.

4.2   Producing SDG indicators according to minimum dimensions of disaggregation

4.3   Examining multi-level disaggregation

To identify those furthest behind, analysis features the simultaneous analysis of disaggregated data on several variables that define the relevant disaggregation dimensions; i.e., multi-level disaggregation. In this Guidebook, this type of analysis is referred to as *multi-level disaggregation analysis or multi-level analysis, for short.*

4.3.1   Example: Poverty rates, Philippines

As countries consolidate their efforts to realize the Sustainable Development Goals related to poverty reduction, it is the responsibility of every sector in the society to be asking questions how they can contribute to building equity in living standards and ensure that the poorest of the poor are not left behind. While a great amount of data is collected about poverty, only a fraction provides the detail that is needed by policymakers to design and assess impact of a well-targeted intervention program on different equity dimensions. With the availability of granular data, socioeconomic planners can capture and use the necessary evidence to understand and respond to inequity in poverty outcomes.

Availability of poverty rate estimates at subnational levels have increased over the years, typically for large geographical administrative areas as illustrated in Figure 4.1 for the 17 *regions*[47] of the Philippines.

Figure 4.1: Poverty Rate by Region, 2018
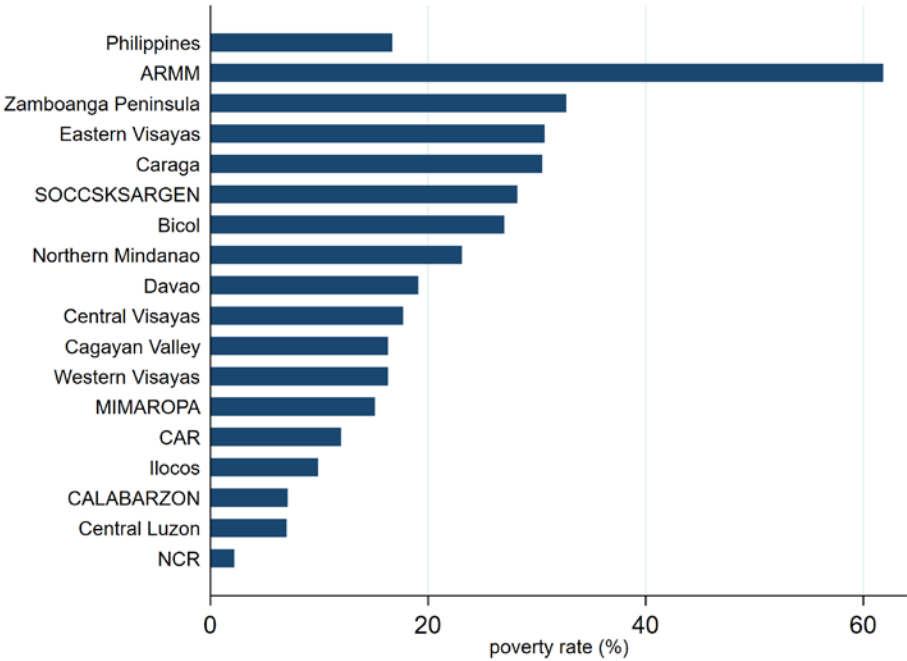
Even at this high level of disaggregation, it is clear that much remains to be done in terms of promoting further geographic equity in poverty reduction as significant pockets of poverty still remain in many areas of southern Philippines like the Autonomous Region of Muslim Mindanao and Zamboanga Peninsula.

Until 2000, reliable data on poverty in specific municipalities of the Philippines were sparse because the Family Income and Expenditure Survey, the survey used to compile official poverty statistics in the country was designed to provide reliable estimates at higher geographic levels. With the efforts of government statisticians and international development partners, the country started compiling municipal- and city-level data on poverty. Based on these statistics, Siayan municipality in the Zamboanga Peninsula was identified as one of the poorest municipalities in the Philippines, with the proportion of the population living below national poverty line estimated to reach as high as 97.5% in 2003 (PSA 2009).

These disaggregated statistics shed light on the prevalence of poverty in Siayan, where majority of its population at risk of socioeconomic vulnerabilities and exclusion. In the succeeding years since these poverty estimates became available, the country's socioeconomic planners, policymakers, and development institutions exerted effort to turn this situation around and improve

---

[47] Regions are the highest levels of geographic disaggregation in the Philippines. Regions are further subdivided into provinces and provinces into municipalities and cities.

the lives of Siayanons. In fact, poverty estimates compiled by the Philippine Statistics Authority showed that the proportion of the municipality's population dropped to 68.4% in 2015. For two consecutive years, Siayan has graduated from the list of 10 poorest municipalities in the country.

In addition to geographically disaggregated poverty data, the Philippine Statistics Authority continues its effort to disaggregate poverty statistics by different dimensions to provide empirical evidence for monitoring poverty and assessing the impact of programs designed to arrest socioeconomic dearth. For instance, to provide inputs for the implementation of the country's *Social Reform and Poverty Alleviation Act* which promotes an area-based, sectoral and focused intervention to poverty alleviation, the PSA has been producing poverty statistics disaggregated by basic sector. Latest statistics show (Figure 4.2) that the prevalence of poverty among farmers, fisherfolks, people living in rural areas, and children is higher than the national average. These results highlight the need for better poverty alleviations programs targeted to these population groups.

Figure 4.2: Poverty Rate by Basic Sector, 2018
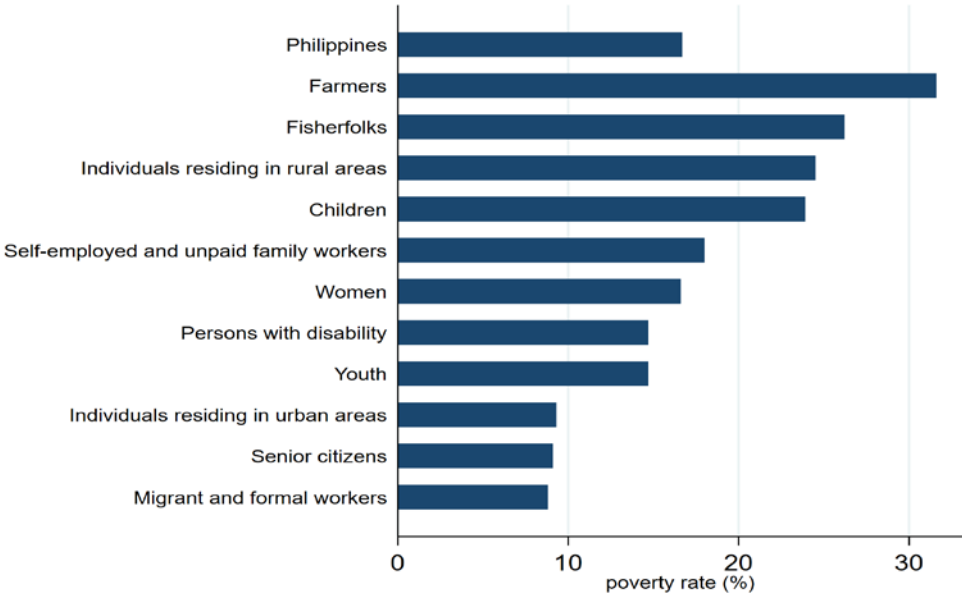
Source:

In summary, availability of disaggregated data on poverty facilitates a more layered and nuanced analysis on equity issues in poverty reduction efforts.

4.3.2   Example: Gender and intersectionality

The United Nations General Assembly adopted by consensus Resolution 70/1: Transforming our world: the 2030 Agenda for Sustainable Development (the 2030 Agenda). The Resolution states that "realizing gender equality and the empowerment of women and girls will make a crucial contribution to progress". The prioritization of gender equality and women's rights is reflected in

a cross-cutting manner throughout the 2030 Agenda, including the declaration, goals, targets and indicators, means of implementation and follow-up and review.

Box 4-1. Briefly-- Measurement of Gender and Intersecting Inequalities

"Inequalities often overlap or 'intersect' in ways that create and compound deprivation and disadvantage. As such, the specific needs of the most marginalized populations need to be identified and measured so they can inform political discourse and spark necessary change.

However, when the data used to inform policies fail to capture such inequalities, decisions are likely to leave people behind. Put differently, if data are incomplete, responses will be incomplete. It is our duty to ensure this does not happen.

Capturing intersecting inequalities requires disaggregating data by sex, age and other characteristics (class, race, location, disability, educational level, migratory status, etc.) It also means carefully deliberating, choosing topics and asking questions about how different groups of people may be differently affected. The process of deliberating and choosing topics, which must involve participation from marginalized groups, is particularly important yet often missing, making data less relevant." *Source: Seck, Papa. 2020. "Integrate intersecting inequalities to leave no one behind" @ https://unstats.un.org/unsd/undataforum/blog/Integrate-intersecting-inequalities-to-leave-no-one-behind/.*

[Counted and Visible: Global Conference on the measurement of gender and intersecting inequalities](https://data.unwomen.org/news/counted-and-visible-global-conference-measurement-gender-and-intersecting-inequalities) @ https://data.unwomen.org/news/counted-and-visible-global-conference-measurement-gender-and-intersecting-inequalities tackled the challenges and demonstrated initiatives on how to measure intersecting inequalities from a gender perspective. The various speakers drawn from government statisticians, policymakers, civil society practitioners, and development organizations shared a wealth of information on: what gender, intersectionality and LNOB means for realizing the SDGs; how is gender, intersectionality and LNOB translated into data production and analysis; a deep dive on vulnerable groups traditionally rendered invisible in social statistics and examples of how data on gender and intersecting inequalities can be strategically used to inform policies and advocacy on gender equality.

Women and girls are often at risk of experiencing disadvantage. Besides sex-based discrimination, they may experience other overlapping forms of discrimination associated with socio-economic characteristics, such as living in rural areas or poor households. These groups of women facing multiple forms of discrimination will experience severe forms of deprivation[48].

The analysis that would be most impactful would involve looking at multiple, contextually relevant, socio-economic characteristics that can be potential causes or drivers of inequality. The rationale is that an individual can be at the intersection of multiple identities that push them behind on the ladder of progress. For instance, the chart in Figure 4.3 below, which summarizes country data on illiteracy rates among those aged 15 to 49, by sex and economic status from 2005-2015, show

---

[48] UN Women, Turning Promises into Action https://www.unwomen.org/-/media/headquarters/attachments/sections/library/publications/2018/sdg-report-gender-equality-in-the-2030-agenda-for-sustainable-development-2018-en.pdf?la=en&vs=4332

gaps not only between the poor and the rich, but also between females and males across countries, mostly to the disadvantage of females among the poorest wealth quintile.

**Figure 4.3: Multilevel analysis of illiteracy rates**

ILLITERACY RATE AMONG POPULATION AGED 15-49, BY SEX AND WEALTH QUINTILES, 2005-2016
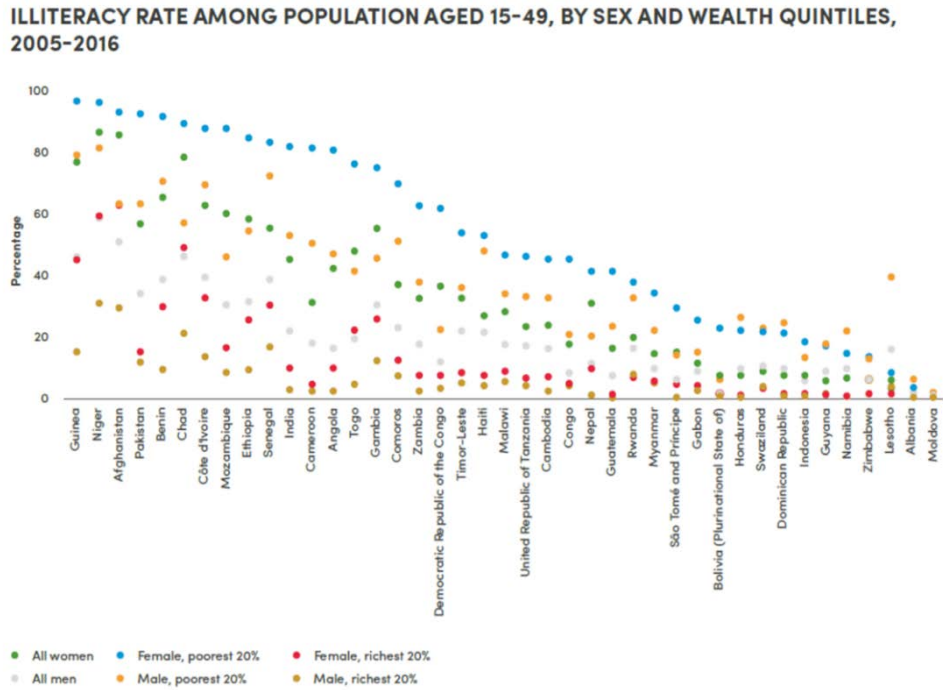
Source: Turning Promises Into Action: Gender Equality in the 2030 Agenda for Sustainable Development (UN Women)[49]

Figure 4.4 below shows that while two-fifths (40.2 per cent) of women aged 18 to 49 in Pakistan were married or in a union before 18 years old, disaggregated data reveals that the gap in corresponding percentages between the poorest in the rural areas (58.5 per cent), and the richest in urban areas (23.9 per cent) is as much as 34.9 percentage points.

---

[49] See https://reliefweb.int/sites/reliefweb.int/files/resources/SDG-report-Gender-equality-in-the-2030-Agenda-for-Sustainable-Development-2018-en.pdf

Figure 4.4: Multilevel analysis of marriage before 18 years old
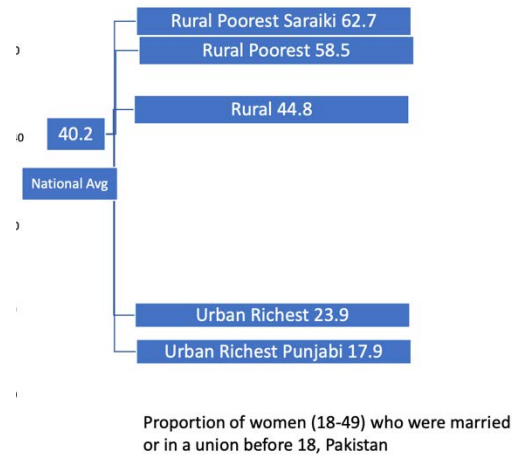
Proportion of women (18-49) who were married
or in a union before 18, Pakistan

Similar patterns of national aggregates masking differences among demographic groups in Nigeria is illustrated in Figure 4.5 below. While nearly half (46.8 per cent) of women in Nigeria aged 18 to 49 were married or in a union before 18, the proportions are much larger (between 8 to 9 out of 10 women) among the poorest in the rural population from the Fulani and Hausa ethic groups, as compared to about a tenth of the Yoruba females among the richest quintile residing in urban areas.

Figure 4.5:

PROPORTION OF WOMEN AGED 18–49 IN NIGERIA MARRIED BEFORE AGE 18, BY LOCATION, WEALTH AND ETHNICITY, 2013
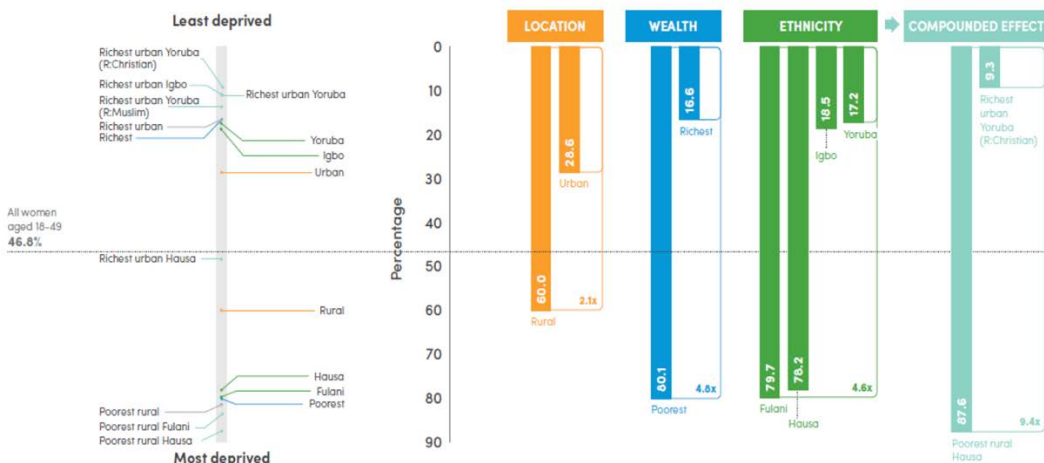
The analyses above can be carried out using a framework for LNOB analysis developed by UN Women. A schematic diagram of the framework is shown in Figure 4.6 below.

Figure 4.6: Framework for LNOB Analysis

Get the right data	•Microdata •Large enough sample size •Sample representative of all groups
Select disaggregati on variables	•Based on previous knowledge of disadvantages •Based on trials for different indicators •Content analysis of policy documents
Conduct analysis	•Calculate SDG indicators (or similar) •Simultaneously disaggregate data using selected variables
Use the results	•Design universal policies that ensure reach to the most disadvantaged •Knock-on effect on SDG achievement •Measure again

Source: UN Women. 2020. Multilevel disaggregation analysis to monitor the SDGs from a Leave No One Behind perspective. https://data.unwomen.org/sites/default/files/documents/Asia-Pacific-Training-Curriculum/Module7/Module7_Syllabus_LNOB.pdf

Recommended reading:
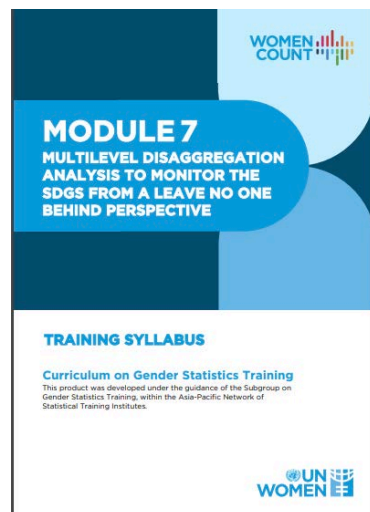
UN Women. 2020. Multilevel disaggregation analysis to monitor the SDGs from a Leave No One Behind perspective https://data.unwomen.org/sites/default/files/documents/Asia-Pacific-Training-Curriculum/Module7/Module7_Syllabus_LNOB.pdf This training module provides the framework for LNOB analysis to carry out multilevel disaggregation analysis on gender and intersectionalities.

---

[51] See https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%208.a_UNWomen_Gender%20data%20and%20multi%20level%20disaggregation.pdf

4.4   Equity assessments and disparity analysis: HEAT

To support countries in measuring and monitoring health inequalities, the World Health Organization (WHO) has developed the Health Equity Assessment Toolkit, a software application that facilitates the assessment of within-country health inequalities. Inequalities can be assessed using disaggregated data and summary measures that are visualized in a variety of interactive graphs, maps and tables. The toolkit is available in two editions:

- HEAT, the Built-In Database Edition, which contains the WHO Health Equity Monitor database, and

- HEAT Plus, the Upload Database Edition, which allows users to upload their own datasets and analyze inequalities using their data.

HEAT, the built-in database edition of the toolkit, uses disaggregated data from the WHO Health Equity Monitor database, which in the 2019 update contained over 30 reproductive, maternal, newborn and child health (RMNCH) indicators, disaggregated by six dimensions of inequality - economic status, education, place of residence and subnational region, as well as age and sex, where applicable. The disaggregated data are based on a re-analysis of over 360 Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS) and Reproductive Health Surveys (RHS) conducted in 112 countries from 1991 to 2017. Over 100 (95 per cent) of countries are low- or middle-income and for 88 countries (79 per cent), data are available for at least two time points.
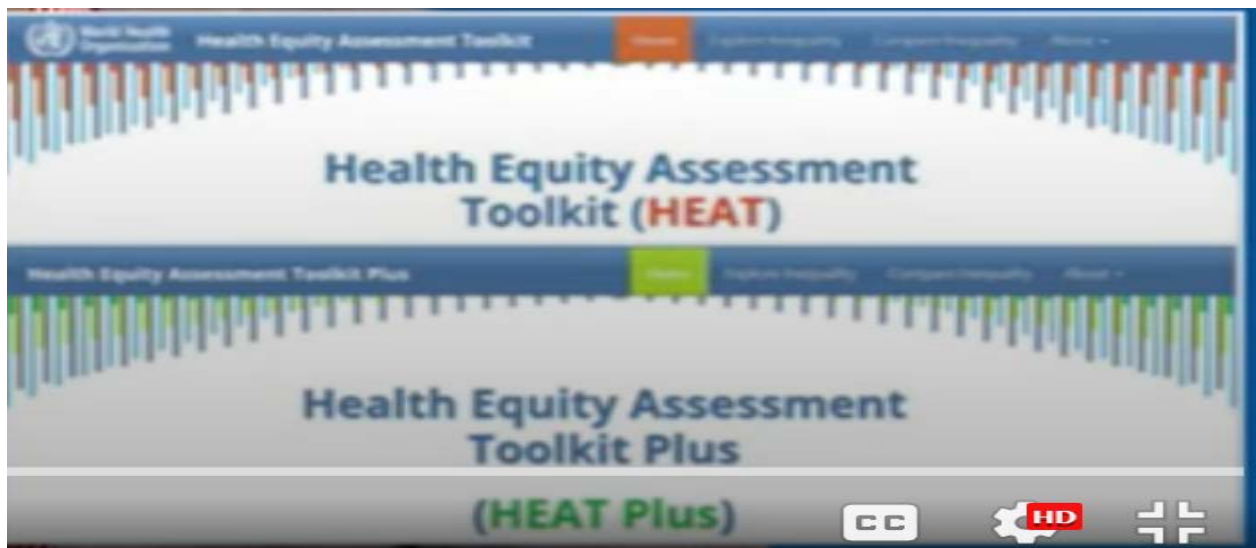
The bar graphs below are taken from the online version of HEAT, pertaining to under-five mortality rates in Bangladesh. Figure 4.7 presents the information disaggregated by sex).

Figure 4.7: Under-5 mortality rate by sex, 1993 – 2014: Bangladesh

Source: https://whoequity.shinyapps.io/HEAT/
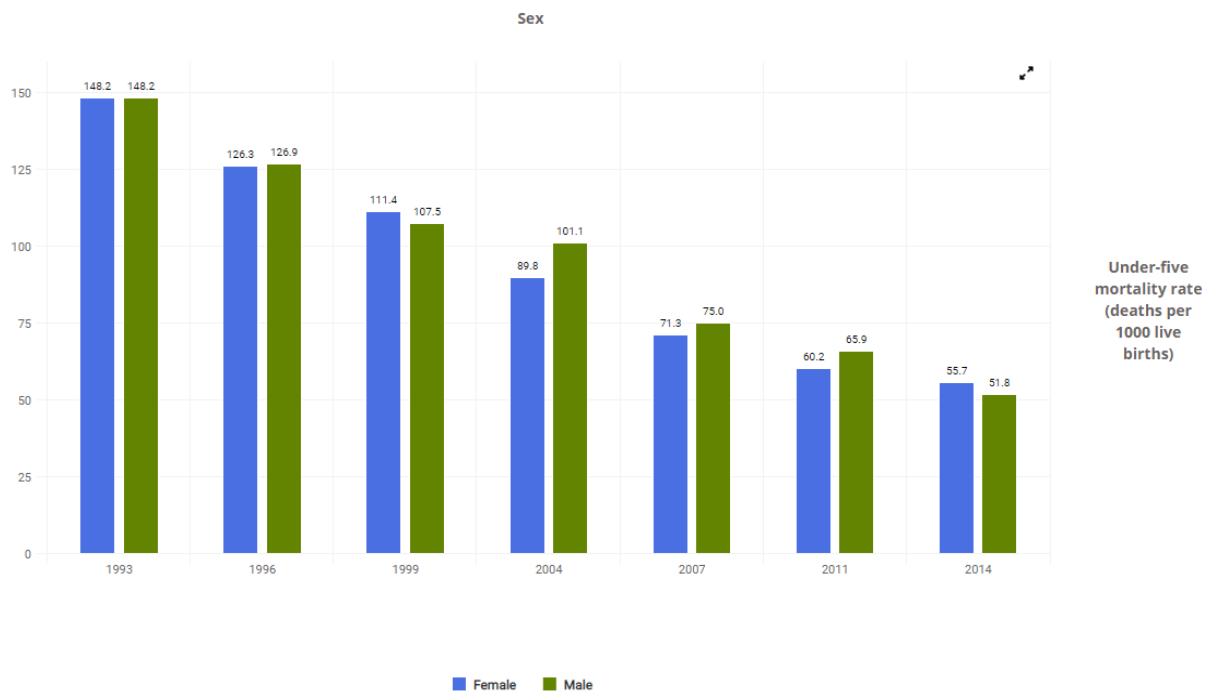
Figure 4.8 shows how child deaths in Bangladesh have dropped significantly across the years from 1993 to 2014. Further, it illustrates that gaps in child deaths between the different wealth quintiles have narrowed considerably over time. In particular, the difference in the number of deaths among children below five between the richest and the poorest quintiles decreased from 87.3 deaths per 1000 live births in 1993 to 24.5 deaths per 1000 live births as of 2014. The ratio of under-five mortality in the poorest to richest quintile ranged from 2.0 in 2011 to 1.7 in 2014.

Figure 4.8: Under-5 mortality rate by economic status, 1993-2014: Bangladesh

Bangladesh (DHS, 1993, 1996, 1999, 2004, 2007, 2011, 2014)

Economic status

Under-five mortality rate (deaths per 1000 live births)

Quintile 1 (poorest) Quintile 2 Quintile 3 Quintile 4 Quintile 5 (richest)

Source: https://whoequity.shinyapps.io/HEAT/
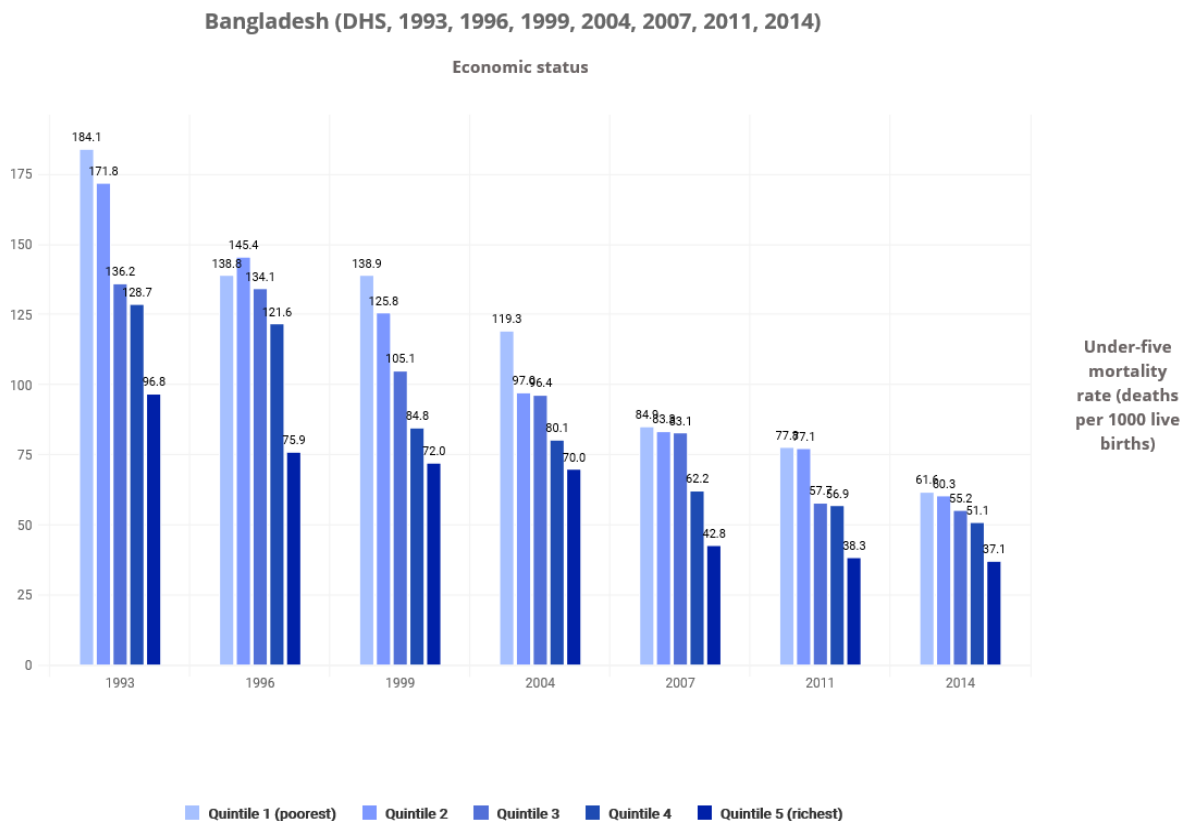
In addition to inequality assessments based on disaggregated data, inequalities may also be assessed using summary measures of inequality. HEAT calculates up to 19 different summary measures of inequality, including both absolute and relative inequality measures. Absolute inequality measures indicate the magnitude of inequality in health between subgroups, while relative inequality measures show the proportional differences in health among subgroups.

For example, Figure 4.8 shows the difference in under-five mortality rates between the richest and poorest quintile in Bangladesh from 1993 to 2014, with a clear pattern of absolute inequality narrowing across time. In contrast, Figure 4.9 illustrates the ratio of under-five mortality rates between the poorest and richest quintiles, showing that relative inequality has fluctuated but changed little over time. This example demonstrates the importance of analyzing both absolute and relative measures of inequality.

Figure 4.9: Ratio of under-five mortality rates poorest and richest quintiles, 1993-2014: Bangladesh

Source: https://whoequity.shinyapps.io/HEAT/
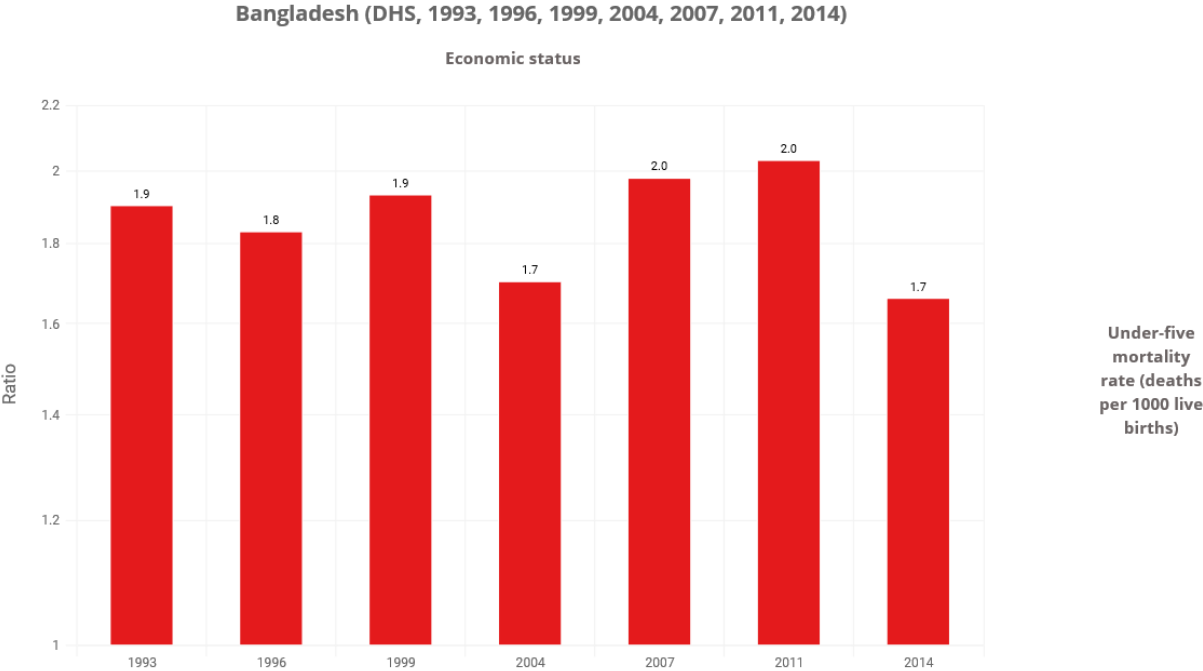
Aside from the difference and ratio measures, which are two simple measures of inequality that make pairwise comparisons between two subgroups, HEAT also calculates complex measures of inequality, which take into account data from all subgroups in order to assess inequalities and provide a more nuanced picture. Absolute complex measures of inequality calculated in HEAT include the absolute concentration index, between-group standard deviation, between-group variance, population attributable risk, slope index of inequality, unweighted and weighted mean difference from best performing subgroup, unweighted and weighted mean difference from mean. Relative complex measures of inequality available in HEAT include the coefficient of variation, mean log deviation, population attributable fraction, relative concentration index, relative index of inequality, Theil index and unweighted and weighted index of disparity. The Handbook on Health Inequality Monitoring published by the WHO discusses in detail the measures used in calculating health inequalities.

HEAT Plus, the upload database edition of the toolkit, has all the same features and functionalities as HEAT and additionally allows users to upload and work with their own datasets. Databases have to be in a specific format in order to be uploaded to HEAT Plus, following the structure laid out in the HEAT Plus template. The flexibility of HEAT Plus makes it a suitable tool for equity assessments at global, national and subnational levels for a wide range of health and health-related indicators and inequality dimensions.

What's Next

Chapter 4 discussed that although household surveys are not well suited to generate spatial disaggregates at lower administrative levels, basic disaggregates such as place of residence, sex, wealth status, and age are usually available. However, innovative approaches such as combinations of these (e.g., urban-poor, young women) or the ability to generate indicators for specific subgroups across many dimensions (e.g., multidimensional poverty) lend more analytical power to the data (Section 4.2).

The next chapter further explains how data visualizations and dashboards can effectively illustrate inequities. Graphs and historical charts in equity analysis tools such as WHO's HEAT are particularly insightful for assessing gaps across disaggregation categories and examining disparities across time. Meanwhile, interactive visualizations and dashboards, such as the SDG Interactive Data Dashboard of a social enterprise known as iTech Mission, show promise in helping citizens obtain snapshots of progress and challenges for attaining the SDGs (Section 5.2.1).
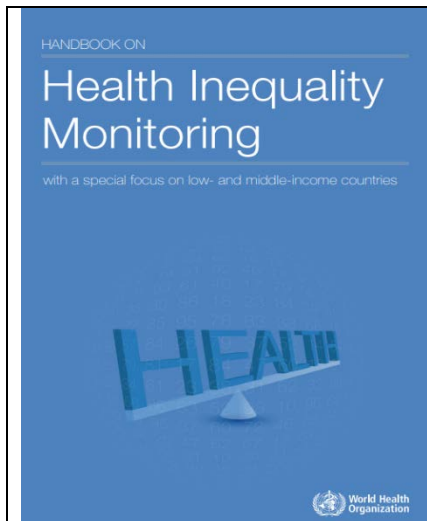
Recommended Reading

WHO Health Equity Monitor (2013–2019).
http://www.who.int/gho/health_equity/en/

It serves as a platform for the Global Health Observatory health inequality database and theme page (containing interactive data visuals, country equity profiles, feature stories, analysis tools and publications). Updated regularly, the Health Equity Monitor is a repository of disaggregated data from over 100 countries (primarily low- and middle-income countries)
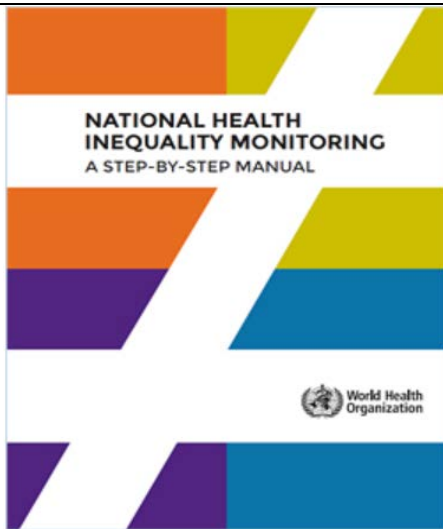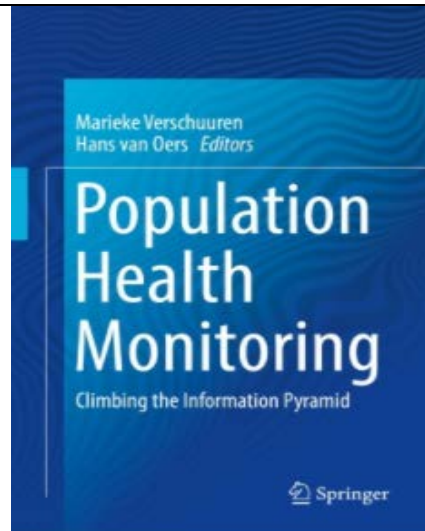

WHO (2017). National health inequality monitoring: a step-by-step manual. Geneva.
http://www.who.int/gho/health_equity/manual/en/

This was designed as a highly accessible, practical reference to encourage and strengthen the practice of health inequality monitoring. The manual is organized according to a flow chart, which shows the steps and sub-steps of the health inequality monitoring cycle, with key questions and itemized checklists of data requirements, analysis/reporting activities and/or decision points. While the manual focuses on health at the national level, the step-by-step approach may be applied to monitor inequalities within any defined population, ranging from a community context to a multinational context.

HANDBOOK ON

Health Inequality Monitoring

with a special focus on low- and middle-income countries

World Health Organization

WHO (2013). Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. Geneva @ https://apps.who.int/iris/bitstream/handle/10665/85345/9789241548632_eng.pdf?sequence=1

This is a user-friendly resource, developed to help countries establish and strengthen health inequality monitoring practices. The handbook elaborates on the components of health inequality monitoring. Throughout the handbook, examples from low and middle-income countries are presented to illustrate how concepts are relevant and applied in real-world situations. It is recognized in the 2014 British Medical Association Medical Book.

Population Health Monitoring Marieke Verschuuren Hans van Oers *Editors* Climbing the Information Pyramid *Springer*	Hosseinpoor AR, Bergen N. Health Inequality Monitoring: A Practical Application of Population Health Monitoring. In: Verschuuren M., van Oers H. (eds) Population Health Monitoring. 2019, Springer, Cham. @ https://link.springer.com/chapter/10.1007/978-3-319-76562-4_8 This chapter draws from contemporary examples to illustrate one application of population monitoring: health inequality monitoring. It demonstrates how monitoring of health inequalities forms a central part of population health initiatives across global and national contexts. The primary aims of the chapter are to demonstrate the importance of health inequality monitoring for equity-oriented policies, programmes and practices; provide examples of how all steps of the health inequality monitoring cycle have been applied; discuss practical challenges of monitoring health inequalities; and suggest strategies for strengthening health inequality monitoring.

References

Haughton, J. and Khandker, S. 2009. Handbook on Poverty and Inequality. Washington, DC., USA: World Bank http://documents.worldbank.org/curated/en/488081468157174849/pdf/483380PUB0Pove101OFFICIAL0USE0ONLY1.pdf (accessed on March 15, 2019)

ICF International. n.d. Demographic and Health Survey (DHS) Programme. Rockville, MD, USA: ICF International. https://dhsprogram.com/ (accessed on March 15, 2019)

ICF International. n.d.. DHS Compiler. Rockville, MD, USA: ICF International. https://statcompiler.com/en/ (accessed on March 15, 2019)

United Nations International Children's Emergency Fund. n.d. Multiple Indicator Cluster Survey (MICS). New York City, NY, USA: UNICEF. http://mics.unicef.org/ (accessed on March 15, 2019)

United Nations International Children's Emergency Fund. n.d.. MICS Compiler. New York City, NY, USA: UNICEF. http://www.micscompiler.org/ (accessed on March 15, 2019)

United Nations Statistics Division (UNSD). n.d.. Global Indicators Database. New York City, NY, USA: UNSD. https://unstats.un.org/sdgs/indicators/database/ (accessed on March 15, 2019)

World Bank. n.d.. Automated Economic Analysis (ADEPT). Washington, DC, USA: World Bank. http://surveys.worldbank.org/adept (accessed on March 15, 2019)

World Bank. n.d.. Living Standard Measurement Study (LSMS). Washington, DC, USA: World Bank. http://surveys.worldbank.org/lsms (accessed on March 15, 2019)

World Health Organization. n.d.. Health Equity Assessment Toolkit (HEAT). Geneva, Switzerland: WHO. https://www.who.int/gho/health_equity/assessment_toolkit/en/ (accessed on March 15, 2019)

WHO eLearning module on health inequality monitoring (2015). https://extranet.who.int/elearn/course/category.php?id=15

World Health Organization. n.d.. Global Health Observatory Data. Geneva, Switzerland: WHO. https://www.who.int/gho/indicator_registry/en/ (accessed on March 15, 2019)

Chapter 5. Reporting, Communicating and Improving Use of Disaggregated Data

Overview

Disaggregating data to meet the LNOB principle is deemed meaningless if not translated to actual policy and program use. Policy advocates can capitalize on available disaggregated data on poverty which provide detailed information on who the poor are, where they live, how many are poor, and why they are poor, to speak on behalf of people who are experiencing severe socioeconomic disadvantage, and make them visible in policy planning. An established SDG and national development plan data reporting system and effective dissemination platforms are thus imperative in ensuring that statistics on those disadvantaged by multiple or intersectional deprivations are accessed by and communicated to the target users, implementers, and the general public.

This chapter highlights the data flows in the reporting of SDG data from national to international repositories and provides examples of indicator dashboards and in-country reporting mechanisms. It also discusses visualizations and dashboards on SDG statistics for ease of conveying patterns of inequities and a case study on how an effective dissemination and communication of SDG statistics result to actual policy uses.

5.1 Reporting on SDG Indicators

5.1.1 Global/regional

Regular preparation of data and updates for the global indicators database on the SDGs requires a complex array of data flows, usually beginning with national data and statistics for global indicators by NSOs and other national data providers. The custodian agencies for the SDG indicators compile internationally comparable data and compute global indicators, informing countries when their data have been adjusted or estimated. Ultimately, the collation and reporting are done by UNSD.

Efficient and transparent transmission of country data, disaggregated data, and metadata for global SDG indicators can be facilitated if NSOs and custodian agencies adopt interoperable statistical data transmission standards, such as the Statistical Data and Metadata eXchange (SDMX)[52]. SDMX can be implemented in incremental steps, and used by non-technicians, but SDMX cannot solve all the data validation challenges countries and custodian agencies face.

Guidelines on data flows and global data reporting for the SDGs have been prepared by the IAEG-SDGs and endorsed by the UN Statistical Commission. These documents can be accessed @ https://unstats.un.org/sdgs/iaeg-sdgs/data-flows/.

As part of its follow-up and review mechanisms, the 2030 Agenda for Sustainable Development encourages member states to "conduct regular and inclusive reviews of progress at the national and sub-national levels, which are country-led and country-driven". These national reviews are

---

[52] See: https://unstats.un.org/sdgs/iaeg-sdgs/sdmx-working-group/

presented a regular reviews by High-Level Political Forum on SDGs (HLPF)[53] These regular reviews by the HLPF are voluntary, state-led, undertaken by both developed and developing countries, and involve multiple stakeholders.

The voluntary national reviews (VNRs) aim to facilitate the sharing of experiences, including successes, challenges and lessons learned, with a view to accelerating the implementation of the 2030 Agenda. The VNRs also seek to strengthen policies and institutions of governments and to mobilize multi-stakeholder support and partnerships for the implementation of the Sustainable Development Goals.

Various guides as well as VNRs prepared by countries since 2018 are available from the website https://sustainabledevelopment.un.org/vnrs/.

5.1.2   In-country

Aside from preparing Voluntary National Reviews (VNRs) of SDG implementation, countries are also disseminating and reporting national statistics on the SDGs through various media. The national SDG indicators do not only include the global SDG indicators available at the country level, but also proxy and supplementary indicators. Data for these national SDG indicators are made publicly available in their respective national reporting platforms (NRP). The NRPs refer to websites, databases, and associated IT infrastructures for collecting, storing, securing, and ultimately disseminating data and related metadata and documentation in an easily accessible format. Target users of the NRPs include government officials and policy makers, members of academic and research institutions, CSOs, and other non-governmental organizations and non-profits, development partners, media and other information providers, the private sector, as well as the general public.

Reporting on SDGs at the country level is dependent on institutional arrangements for statistical coordination, the legal framework that involves the sharing of data, and the understanding that data are a public good. A significant challenge, whether for country or global monitoring of the SDGs, is the availability of disaggregated data. The 2017 ADB-ESCAP survey conducted in 22 countries revealed that among reporting economies, there is considerable disaggregation of statistics by location for many SDG indicators. However, data disaggregation is sparse for some indicators by sex, and even further limited —if not absent—for persons with disability, and indigenous peoples. The 2018 report of ADB and UN Women on "Gender Equality and the Sustainable Development Goals in Asia and the Pacific" also pointed out that of the 85 unique gender-related SDG indicators they examined, only a quarter (26 per cent) are available for more than two-thirds of the countries or territories in the Asia-Pacific region, while 41 per cent have no relevant regional data.

5.2   Communicating Data and Statistics on SDGs

5.2.1   Data Visualizations and Dashboards for Effectively Illustrating Inequities

---

[53] More information on the HLPF is available on the website https://sustainabledevelopment.un.org/hlpf

An effective way to communicate data and statistics is through storytelling, made more meaningful with the use of visualizations. Visualizations of disaggregated data the communication of inequalities across disaggregation dimensions and categories. They can make patterns easier to see or they can reveal patterns that might otherwise have been hidden.

When properly designed, visual information on disaggregation dimensions enables data exploration and effective storytelling that informs decision-making. The communication of data and statistics, including data visualization and storytelling enables us to uncover, understand, and communicate patterns in disaggregated data, especially for those in multiple or intersectional deprivations (e.g. "the proportion of working children in the poorest quintile in urban areas"; "the proportion of children with access to safe water among the poorest quintile in rural areas").

For example, the spider plot (also called radar chart) in Figure 5.1 displays several indicators on reproductive health, education, decision-making, and access to finance in the form of a two-dimensional chart with the disaggregated data for the poorest rural Indian women and the corresponding richest Indian women in urban areas to illustrate inequalities.

Figure 5.1: Spider chart visualization of multidimensional Inequalities

INEQUALITIES BETWEEN POOREST RURAL AND RICHEST URBAN INDIAN WOMEN, VARIOUS INDICATORS, PERCENTAGE, 2015-2016
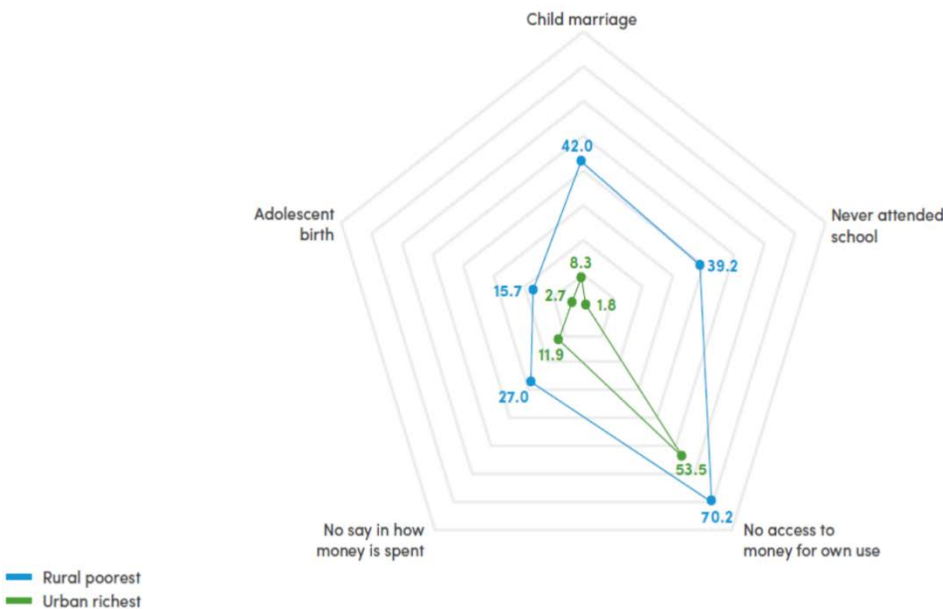
Source: UN Women calculations based on microdata from the India National Family Health Survey (NFHS-4/DHS).

Graphs and historical charts in equity analysis tools such as WHO's HEAT described in Chapter 4 are particularly insightful for assessing gaps across disaggregation categories and examining disparities across time.

Interactive visualizations and dashboards, such as the SDG Interactive Data Dashboard (Figure 5-2 shows a screenshot) of a social enterprise known as iTech Mission, show promise in helping citizens obtain snapshots of progress and challenges for attaining the SDGs. The iTech Mission

tool makes use of data from the UN Statistics Division's SDG Global Database, in a way that allows its users to explore and visualize data for analyzing SDG progress. Users can assess their country's progress on the SDGs, and explore trends on a particular goal, target, or indicator using graphs, bubble or rank charts, tree maps or tables. In addition, users can select and display SDG indicators using different metrics. A monitoring tool displays a given target and its indicators, and provides information on the data value, unit and trend. Users can also view trends over time towards the target or choose to track and monitor data on the SDGs that are most important to them, or create country profiles on the SDGs. The Dashboard allows linking to the data from NSOs and other government databases. A 'Citizens Vote' tool is also available which allows users to identify six (out of 16) issues that are most important to them, ranging from a good education to action taken on climate change, and to see how their priorities compare with others across the world. The votes are recorded with the user's sex, age, education level and location.

Figure 5.2: Features of the SDG Interactive Data Dashboard

Source: https://www.sdgsdashboard.org/
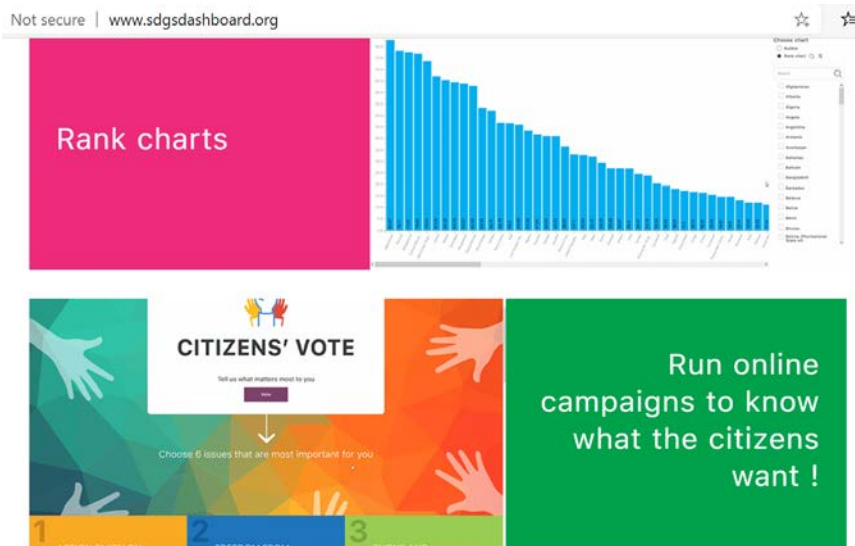
Dashboards are linked to databases with the ability to pull real-time data from multiple sources and are designed to provide an at-a-glance view of vast amounts of information synthesized through graphs, indicators, symbols, and others.

Countries are beginning to experiment on using SDG dashboards, with varying purposes and success. Table 5.1 provides a list of data portals linked to the UNSD-FCDO Project on SDG Monitoring[54].

Table 5.1:

Country	Link/info
Bangladesh	Online: http://www.sdg.gov.bd/
Burundi	Online: https://burundi.opendataforafrica.org/addin/sdg
Cambodia	Online: http://camstat.nis.gov.kh/#/
Ghana	Online: https://sustainabledevelopment-ghana.github.io/
Jordan	Online: Front-end: http://dosweb.dos.gov.jo/sdgs/ Back-end: http://jorinfo.dos.gov.jo/Databank/pxweb/en/SDG/
Kyrgyzstan	Online: https://sustainabledevelopment-kyrgyzstan.github.io/
Mozambique	Partially online: https://mozambique.opendataforafrica.org/sdg
Palestine	Online: https://www.pcbs.gov.ps/mainsdgs.aspx
Rwanda	Online: https://sustainabledevelopment-rwanda.github.io/
Tajikistan	Work in progress to develop their own solution as part of the MoU

---

[54] Refer to https://unstats.un.org/capacity-development/UNSD-FCDO/

Country	Link/info
Uganda	Online: https://uganda.opendataforafrica.org/sdg
Zambia	Online: http://zambia.opendataforafrica.org/addin/sdg
Zimbabwe	Online: https://zimbabwe.opendataforafrica.org/sdg

5.3 Improving use of results of multi-level analysis

5.3.1 How effective dissemination and communication leads to policy uses of disaggregated statistics: A case study

Access to granular data on poverty has a wide range of uses for socioeconomic planners and policymakers. Granular data on poverty which provide detailed information on who the poor are, where they live, how many are poor, and why they are poor, facilitate formulation of more efficient and effective strategies and programs on poverty reduction such as social protection programs, including conditional and non-conditional cash transfers, provision of unemployment benefits, employment facilitation, tax reforms, among others. Granular data on poverty can also be used to assess the impact of these programs over time.

The Philippine experience in collecting granular data on poverty is a good example showcasing a wide range of policy uses. The Philippine Statistics Authority conducts the triennial Family Income and Expenditure Survey, which is the main source of official poverty statistics. Until recently, FIES was designed to provide reliable estimates at the regional level. However, due to demand for more geographically disaggregated poverty statistics, starting in 2003, provincial poverty statistics have also been published, with caveat for provinces with large sampling error.[55] In 2005, the Philippine national statistical system collaborated with the World Bank to undertake a poverty mapping project which resulted to the generation of municipality- and city-level poverty statistics using small area estimation techniques. Since then, the PSA has updated the small area poverty estimates with the availability of new rounds of FIES and Census of Population and Housing.

Since municipality- and city-level poverty statistics became available in 2005, national and local government units (LGUs) in the Philippines have used the data as inputs for formulating and implementing poverty reduction programs. For example, the Philippine Department of Social Welfare and Development (DSWD) had used the estimates to identify poor municipalities for its National Household Targeting System for Poverty Reduction (NHTS-PR) data collection; Kalahi-CIDDS Program; assistance for families affected by typhoon Yolanda (Haiyan) in Western Visayas; Student Grants-in-Aid Program for Poverty Alleviation in the Cordillera Administrative Region; and Cash-for-Training Program in the Soccsksargen Region. The LGUs of Negros Occidental, Pangasinan, municipality of Nabas in Aklan, city of Baguio, and municipalities of La Trinidad, Itogon, Sablan and Tuba in Benguet, also refer to the SAE as inputs in developing local

---

[55] Starting 2018 FIES, the survey adopted the 2013 Master Sample Design, with a sample size of approximately 180,000 sample households (i.e., four times the previous sample size), which significantly improved the reliability of provincial poverty statistics and allowed the generation of poverty statistics at the highly urbanized city leve (PSA 2019).

socioeconomic profiles and situationers for identifying areas requiring poverty intervention programs The provincial governments of Aklan, La Union, Negros Occidental, Pangasinan and Southern Leyte use the estimates in assessing the implementations of their poverty reduction programs. Further, international development organizations also use small area poverty estimates in various poverty-related alleviation programs in different areas in the country (PSA 2013 and 2016).

In addition to the official poverty statistics (national, regional, and provincial) compiled by PSA using data from FIES and municipal- / city-level poverty statistics compiled using small area estimation technique, other government agencies also have its own initiatives to collect granular data on poverty. For instance, the **Community-Based Monitoring System (CBMS),** which was originally designed to provide policymakers and economic planners with a reliable information base that can be used for tracking the impacts of macroeconomic reforms and various policy shocks in the early 1990s, has evolved to include collection of disaggregated data that can be used for planning, program formulation, policy impact and poverty monitoring. Since 2000, CBMS has been adopted by local government units, and various government agencies have recognized the usefulness of CBMS for various thematic concerns such as local planning, grassroots participatory budgeting, poverty diagnosis, monitoring the Millennium Development Goals (MDG), disaster-risk reduction management (DRRM) and climate change adaptation (CCA), gender and development, impact monitoring, and food insecurity among others. In 2019, the Philippine Government enacted a law to institutionalize the use of CBMS as a tool for poverty alleviation program formulation and implementation.

Furthermore, the National Household Targeting Office (NHTO) of DSWD manages the National Household Targeting System for Poverty Reduction (NHTS-PR) or *Listahanan*. The database identifies poor families, through proxy means test to estimate household income based on various social, economic, and housing characteristics. The database is accessible to national government agencies and other social protection stakeholders for identifying potential beneficiaries of social protection programs. Availability of such database aims to minimize poorly targeted social services resulting from exclusions of the poor from necessary social services, therefore, minimizing wastage of resources on those who are not actually poor. Recently, when the Philippine Government launched the Social Amelioration Program in response to the pandemic brought by COVID-19, NHTS-PR were used to identify the poorest of the poor as target beneficiaries to receive cash grants.

5.3.2   Open data principles increase access and, hence use of disaggregated data

Making use of aggregated and disaggregated data to inform policy and social good starts with open data principles, that is, having data that are freely available online for anyone to use and republish for any purpose. Aggregated (and disaggregated) data are typically the final products of NSOs and other data producers in a national statistical system; these are used to monitor trends in development and to inform public policy. Traditionally NSOs have disseminated aggregated and disaggregated data through yearbooks and publications (including census and sample survey reports). But with the rapid growth of the Internet, most NSOs disseminate data and statistics on their websites. Further, using survey cataloguing software such as the World Bank's National Data Archive (NADA) platform, several NSOs have also made microdata —

survey responses excluding identities of respondents to protect data privacy and confidentiality—available for public use. Microdata of surveys supported by the development community, such as DHS, MICS, and the LSMS, are also readily available for public use. Data from DHS and MICS are available for exploration, may be aggregated and disaggregated at different levels for examination over time and for cross-country comparison through platforms like StatCompiler and MICS Compiler. When microdata of LSMS or other surveys that provide information on poverty are available, they can also be examined more systematically for generating aggregated and disaggregated statistics on poverty and inequality with the World Bank's software platform called ADePT.

What's Next

Following the IAEG-SDGS' work program in 2020 and the commitment made during the 51st session of the United Nations Statistical Commission (UNSC) to proceed with the data disaggregation work to meet the LNOB principle, the group designed a stock-taking questionnaire to compile an inventory of existing methodologies, guidelines, and tools specific for data disaggregation that have already been developed and agreed by key expert/international groups, mechanisms and committees [56]. This also comprises an inventory of tools or guidance for reporting and dissemination of disaggregated data.
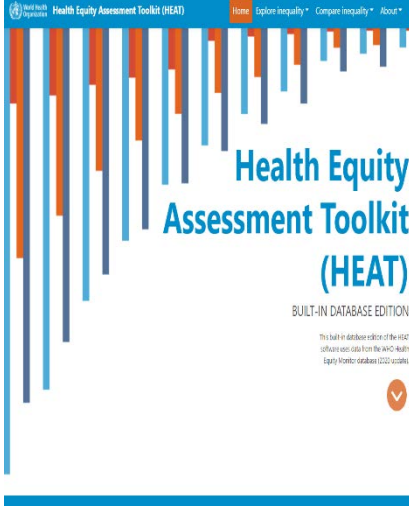
Some of the good practices and guidelines identified in the inventory for the dissemination of disaggregated data of select focused population groups include:

- On refugee and internally displaced people (IDP) statistics, a compiler's manual is available to provide a clear guidance on the production, promotion, and managements of dissemination products. A Joint IDP Profiling Service (JIPS) Essential Toolkit gives an overview on validation, reporting and dissemination and emphasizes the relevance of data dissemination for preventing duplication, increasing transparency and use. Good practices include uploading data on the JIPS Dynamic Analysis & Reporting Tool (DART) and the Humanitarian Data Exchange (HDX). JIPS has also established jointly with partners a story map on Progress Towards Durable Solutions in North Darfur, Sudan as an effective dissemination tool.

- On disability statistics, the WHO supports countries conducting the Model Disability Survey in analyzing and reporting the data which will inform policy makers on what actions need to be taken. A Disability Data Portal is also accessible to provide a snapshot of the available global data on people with disabilities in 40 countries. The portal shows key development indicators mostly sourced from the SDGs as well as other indicators relevant to the UN Convention on the Rights of Persons with Disabilities. On food insecurity and related statistics, the FAO reports that a guidance material on communicating the

---

[56] The participants to the stock-taking activity included established city and expert groups under the UNSC, international expert groups and committees, as well as UN agencies and stakeholders focusing on vulnerable populations, disaggregation dimensions and data collection methods (i.e., censuses, surveys, administrative sources, etc.). The results of the questionnaire are provided in a summary document on compilation of tools and guidelines to provide countries with a resource platform on data disaggregation. The summary document is intended to be regularly updated and will eventually be made accessible online through the following website: https://unstats.un.org/sdgs/.

accuracy measures/error profiles was developed and tested to provide users the freedom to decide for themselves whether the disaggregated estimate are reliable for use (UN 2020).

Recommended Reading

Health Equity Assessment Toolkit (HEAT): software for exploring and comparing health inequalities in countries. http://www.who.int/gho/health_equity/assessment_toolkit/en/

It is a software package that that facilitates the assessment of within-country health inequalities. Users can create customized visuals based on disaggregated data or summary measures. There are two editions of the toolkit: HEAT, Built-In Database Edition, which includes the Health Equity Monitor database; and HEAT Plus, Upload Database Edition, which allows users to upload and work with their own database.

iTech Mission. n.d.. SDGs Dashboard. http://www.sdgsdashboard.org

This chapter draws from contemporary examples to illustrate one application of population monitoring: health inequality monitoring. It demonstrates how monitoring of health inequalities forms a central part of population health initiatives across global and national contexts. The primary aims of the chapter are to demonstrate the importance of health inequality monitoring for equity-oriented policies, programmes and practices; provide examples of how all steps of the health inequality monitoring cycle have been applied; discuss practical challenges of monitoring health inequalities; and suggest strategies for strengthening health inequality monitoring.

References

ADB and UN Women. 2018. Gender Equality and the Sustainable Development Goals in Asia and the Pacific: Baseline and Pathways for Transformative Change by 2030. Bangkok, Thailand: ADB and UN Women. https://www.adb.org/publications/gender-equality-sdgs-asia-pacific (accessed on March 15, 2019)

Hosseinpoor AR, Nambiar D., Tawilah, J., Schlotheuber A., Briot B., Bateman M., Kusumawardani N., Myint T., Nuryetty MT., Prasetyo S., Floranita S&F. Capacity building for health inequality monitoring in Indonesia: enhancing the equity orientation of country health information systems. Global Health Action. 2018, Taylor & Francis Online. https://doi.org/10.1080/16549716.2017.1419739

UN Development Programme (UNDP). 2017. SDG Dashboards The role of information tools in the implementation of the 2030 Agenda http://www.asia-pacific.undp.org/content/dam/rbap/docs/meetTheSDGs/SDG%20Dashboards%20UNDP-SIGOB.pdf

UNECE. n.d.. "Making Data Meaningful" guide series- Parts 1, 2 and 3. https://unstats.un.org/unsd/EconStatKB/KnowledgebaseArticle10350.aspx (accessed on March 15, 2019)

WHO eLearning module on health inequality monitoring (2015). https://extranet.who.int/elearn/course/category.php?id=15

WHO (2017). State of Health Inequality: Indonesia. Geneva. https://www.who.int/gho/health_equity/report_2017_indonesia/en/

Chapter 6. Producing and Using Disaggregated Data: From Potential to Possible

Overview of Chapter

This Guidebook provides information on existing statistical sources, methods, tools and current initiatives that address some of the key issues that need to be considered in the production and analysis of data needed for generating disaggregated statistics and indicators and the reporting and communication of such. This final chapter looks at key issues on why required disaggregated data are often not available and provides information on knowledge products and ongoing capacity development initiatives that may help national statistical systems to address these issues.

6-1 Barriers to producing and using disaggregated data

Target 17.18 of the SDGs strongly and urgently recognizes that data and statistics, particularly disaggregated data, are crucial means of implementation for the goals—

> *By 2020, enhance capacity-building support to developing countries, including for least developed countries and small island developing States, to increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts*

Available disaggregated data fall short of the data needs of the 2030 Agenda-- , especially due to lack of resources and capacity. These challenges present opportunities for innovation to ensure that the SDGs become truly transformative and reach those furthest behind. Countries and the development community have a shared responsibility to ensure that NSSs are effective and capable of producing high-quality disaggregated data for monitoring the SDGs. This requires investments in human resources, new technologies, infrastructure, data architectures, geospatial data and management systems, as well as information intermediaries. Capacities for communicating data effectively must be developed to promote opportunities in improving data use that will contribute to ultimately attaining sustainable development.

A 2018 survey shows that areas where support is needed for data disaggregation vary across regions (and countries), as summarized in Figure 6-1.

Figure 6-1: Priorities for data disaggregation

Supporting the monitoring of SDGs in developing countries: Priorities for data disaggregation

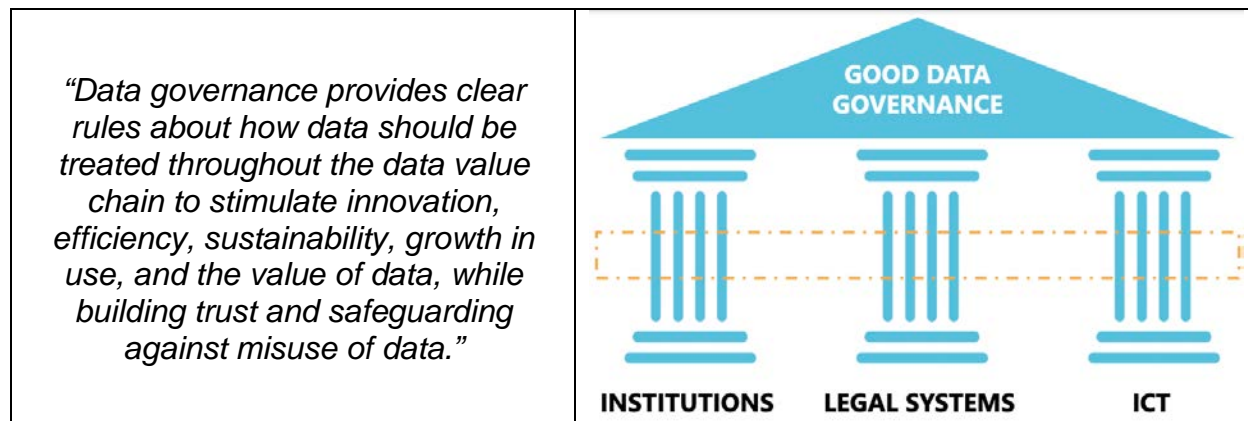*Please indicate what types of data disaggregation require the most immediate support.*
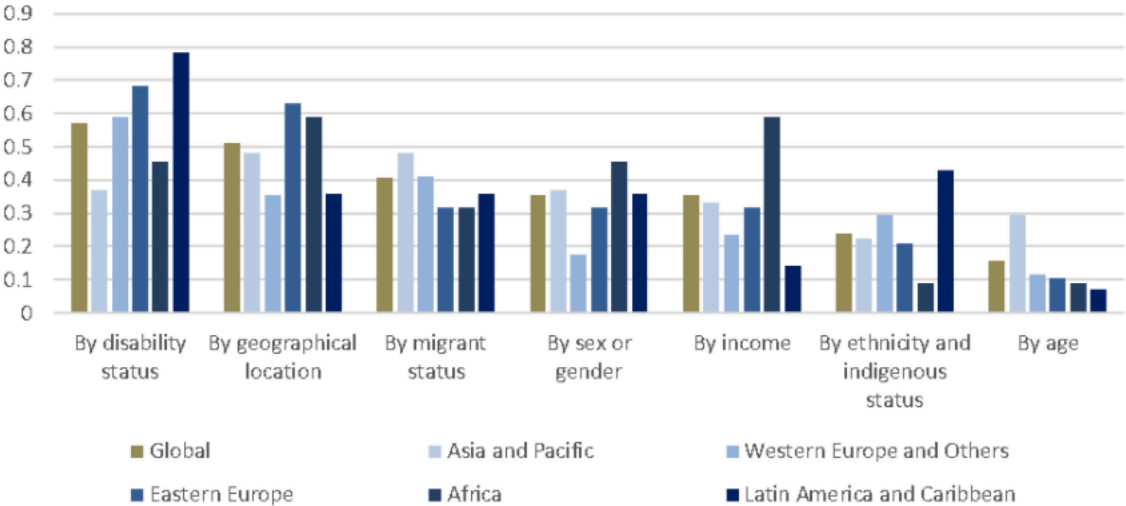
Source: HLG-PCCB/PARIS21 (2018). *Survey results: new approaches to capacity development and future priorities.* Paris: PARIS21. Retrieved from https://paris21.org/capacity-development-40/cd40-survey.

Apart from the technical challenges illustrated in the previous chapters, data governance issues relating to the human rights principle "do no harm" need to be addressed. That is, data collection exercises should not create or reinforce discrimination, bias or stereotypes against population groups. Privacy, confidentiality and data protection must be ensured particularly for digit-based data (big data, GIS) as well the more disaggregated or granular data collected and the sometimes small vulnerable populations where being visible may be harmful at the individual level.

A presentation on "Gender data governance and official statistics"[57] is a good introduction to the issues.

---

[57] UN Women, UNSD, and Intersecretariat Working Group on Household Surveys (ISWGHS). Global conference on the measurement of gender equality, leave no one behind and intersecting inequalities: Concept Note and Agenda. New York, USA. 26-27 February 2020.
https://data.unwomen.org/sites/default/files/documents/gender%20and%20intersecting%20inequalities/docs/presentations/2.8.1.%20Open%20data%20watch.pdf

"Data governance provides clear rules about how data should be treated throughout the data value chain to stimulate innovation, efficiency, sustainability, growth in use, and the value of data, while building trust and safeguarding against misuse of data."

Source:
https://data.unwomen.org/sites/default/files/documents/gender%20and%20intersecting%20in equalities/docs/presentations/2.8.1.%20Open%20data%20watch.pdf

6-2 Areas where capacity development are needed

NSOs and development partners are exploring potential synergies and opportunities for collaboration to strengthen the data for development ecosystem. When using and integrating innovative data sources with traditional data sources in the statistical production system, NSOs are proceeding in steps, starting with clearly stating the focus of the analysis, consulting with data science experts, and obtaining lessons from experiences of other NSOs. As such initiatives involve risks and experimentation, NSOs recognize the need for flexibility and patience in evaluating the potential for these innovations for ultimately meeting the disaggregated data requirements of the SDGs on leaving no one behind.

Indonesia's approach in improving their health information system is highlighted in Box 6-1.

Box 6-1. Enhancing the equity orientation of health information systems: Indonesia
Indonesia enhanced the equity orientation of its health information system through capacity building activities undertaken by WHO between April 2016 and December 2017. This involved a series of workshops, meetings, and processes on health inequality monitoring including identification of health topics/areas of interest, mapping data sources and identifying gaps, conducting equity analyses using raw datasets, and interpreting and reporting inequality results. As a result of these activities, WHO and Indonesia's Ministry of Health jointly produced the first national report on the state of health inequality. Sustained political will across levels of administration and leadership is the main reason for the progress of health inequality monitoring in the country. Challenges experienced during the process highlighted the need for the availability of raw datasets, active collaboration across various government units, further capacity building in quantitative and qualitative research methods, as well as routine inclusion of dimensions of inequality in surveys, civil registration, health facilities and other relevant data sources through a consensus across arms of government to include priority dimensions of inequality. This capacity building for health inequality monitoring exercise in Indonesia is adaptable to other contexts.

Zeroing in on: Big Data

Making use of innovative data sources involves technological infrastructure, both hardware and software. Data analysis software tools may not be suitable or efficiently used for large datasets in a sequential computer. NSOs need improved information communication and technology (ICT) infrastructure to download big data (bandwidth), as well as to catalogue, organize, and process big data in a sufficiently timely manner. The availability of interfaces by some statistical packages such as open sourced R to Hadoop – MapReduce for most used statistical platforms, however, significantly contributed to the use of big data analytics. A related technological issue is the curation of data: since big data results in a messy collage of data points whose accuracy is challenging to establish. On new skill sets, while NSOs have experience in curating data, many have no data scientists that are strong in both data and computational focus. NSOs and other data producers are also recognizing the need for new legal protocols and institutional arrangements to access big data holdings for development purposes, as well as to prevent the misuse of big data.

While big data has been customarily characterized to have 3 Vs: (large) volume, velocity and variety, big data is not just about the data, the "crumbs" or digital exhaust, but also about capacities including econometric tools, software, and hardware for uncovering patterns and the community or data ecosystem. Harnessing data from innovative data sources such as geospatial information, earth observations, mobile data, social media and/or crowd-sourced data with those from traditional data sources requires data interoperability[58], the ability to access and process data assets from multiple sources and multiple formats without losing meaning and then integrate these assets into coherent information products or services (e.g., for mapping, visualization, and other forms of analysis).

Readiness for use of big data in official statistics
https://unstats.un.org/bigdata/task-teams/training/UN_BigData_report_v5.0.html

Barriers

Key point 1 - Strategic coordination: Only a third of all NSOs have overarching big data strategies in place and Chief Data Officers only exist in some NSOs. The biggest challenge for NSOs is the collaboration with Big Data source owners outside the government.

Key point 2 – Legal framework: Legal frameworks are still insufficient to regulate big data applications. Only a small share of NSOs rely on legal frameworks that guarantee access to big data.

Key point 3 – IT infrastructure: IT infrastructure appears as central barrier to develop big data capacity; onsite and offsite storage capacity needs improvement for many. Only a few NSOs consider cloud storage a relevant option.

---

[58] Refer to: Luis González Morales & Tom Orrell. 2019. Data interoperability: a practitioner's guide to joining up data in the development sector @
http://www.data4sdgs.org/sites/default/files/services_files/Interoperability%20-%20A%20practitioner%E2%80%99s%20guide%20to%20joining-up%20data%20in%20the%20development%20sector.pdf

Key point 4 – Human Resources: Most NSOs lack a competency framework to develop new skills to cope with big data (mobile phone, geospatial data) and new methodologies (machine learning).
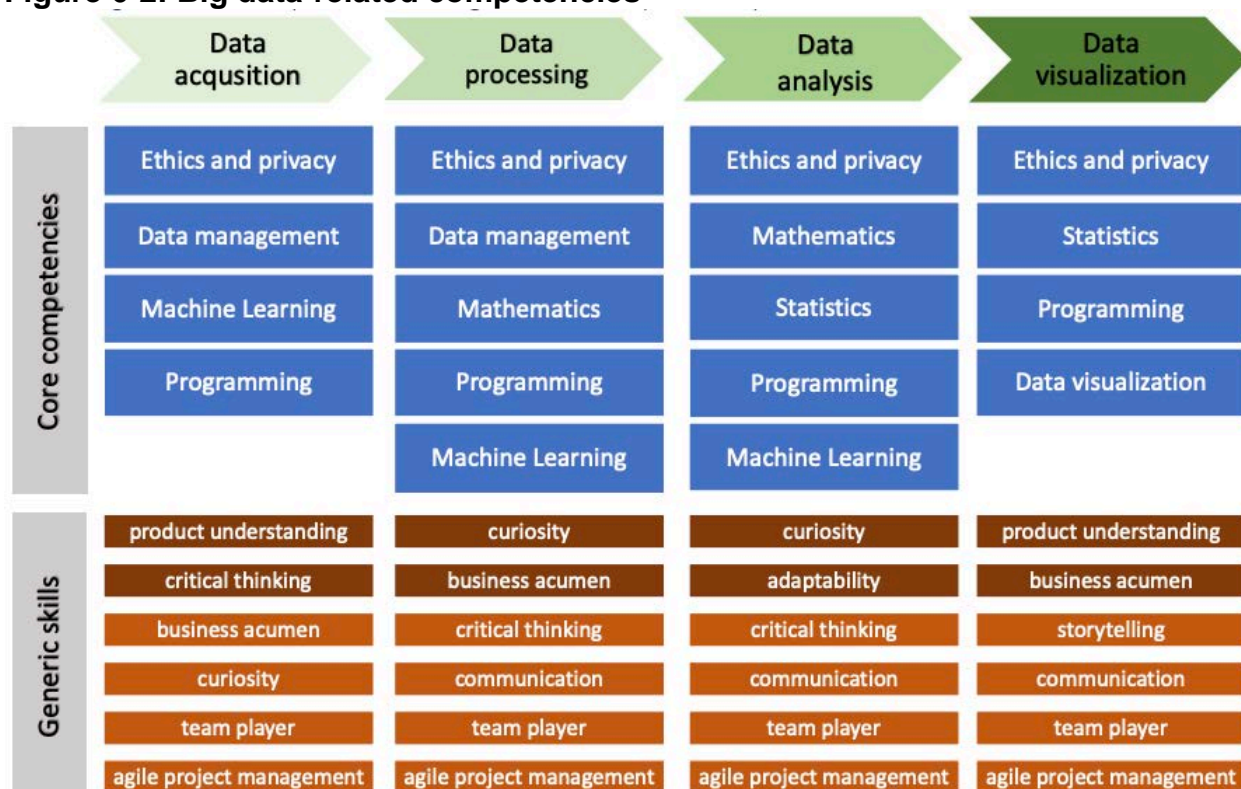
Training on Big Data

https://unstats.un.org/bigdata/task-teams/training/index.cshtml

The above-stated circumstances impose immense pressure on official statistics. On the one hand, its role to ensure the highest standards and quality of statistical information becomes even more vital in the era of fake news and post-truth. On the other hand, it is expected to keep up with the growing demands of data users. To this end, attempts to modernize statistical production have been increasingly undertaken by national statistical organizations (NSOs), since they recognize the potential of novel data processing techniques and new data sources. Among the latter, big data have been of particular interest to the NSOs. Yet, they entail another challenge, not only at the level of their acquisition and implementation into the statistical production, but also in the realm of sustaining relevant skills which reach beyond the traditional set of statistical competencies.

To address this challenge, the UN Global Working Group Task Team on Training, Competencies and Capacity Development has developed this Competency Framework for use by NSOs. It covers the wide array of skills and knowledge considered relevant for those working with big data acquisition and processing. The proposed framework involves core competencies, as well as a more general set of soft skills. The framework is summarized in Figure 6-2.

Figure 6-2: Big data-related competencies

Source: https://unstats.un.org/bigdata/task-teams/training/UNGWG_Competency_Framework.pdf

6-3   Investing in disaggregated data

> **Box 6-2.** *What this means: Statistics Canada*
> In 2018/2019, the government allocated $6.7 million in funding over five years, with $600,000 thereafter, to create the Centre for Gender, Diversity and Inclusion Statistics (CGDIS). The goal of the CGDIS is to support evidence-based policy and program development by monitoring and reporting on gender, diversity and inclusion. In addition to this funding, Statistics Canada received $4.2 million in 2019/2020 for four distinct activities, as part of Canada's Anti-Racism Strategy.
> Statistics Canada continues to work with its federal, provincial and territorial counterparts to develop ways to paint a more complete picture of the situation.
> For more information: *https://www.statcan.gc.ca/eng/transparency-accountability/disaggregated-data*

Despite the importance of having disaggregated information, there are several reasons why official statistics are usually expressed as aggregated data. Budget constraints are foremost. In some cases, the vulnerable groups are hard to capture in conventional data collection systems. For instance, people living on the streets may not necessarily be covered in household surveys because they are not included in sampling frames. It is also difficult to get data from people living in very remote and/or conflict areas. Even availability of financial resources to collect data on these groups will not suffice. Additionally, in cases where resources might be sufficient, many policy makers are more comfortable with tracking a limited number of aggregate figures rather than studying a more complex array of information disaggregated across multiple dimensions. For this reason, the development of new methods and increased affordability of statistical disaggregation tools must be accompanied by new techniques for communication of these statistics via dashboards and other communication tools so that non-statisticians can grasp their meaning and importance for policy formulation, implementation and monitoring.

Global Initiatives: The CTGAP

A systematic way of building capacity and mobilizing the needed resources to do so has been mapped out in the Cape Town Global Action Plan for Sustainable Development Data (CTGAP)[59]. Of particular relevance to disaggregated data is Objective 3.5 of the Plan:

---

[59] Reference: https://unstats.un.org/sdgs/hlg/Cape-Town-Global-Action-Plan/

The Plan also highlights the importance of new technologies and data sources that have been discussed in this Guidebook:

With respect to new data sources and other data, *citizen-generated data (CGD)* as a source is increasingly being explored by national statistical systems. A plethora of concepts and initiatives use CGD to achieve many goals, ranging from citizen science, citizen sensing, and environmental monitoring to participatory mapping, community-based monitoring, and community policing. In these initiatives, citizens may play very different roles - from taking on the role of mere sensors, to giving them agency to shape what data gets collected. Initiatives may differ in respect to the media and technologies used to collect data, the ways stakeholders are engaged with partners from government

or business, and in terms of how activities are governed to align interests between these parties[60].

Figure 6-3: What is CGD

Source: https://paris21.org/cgd
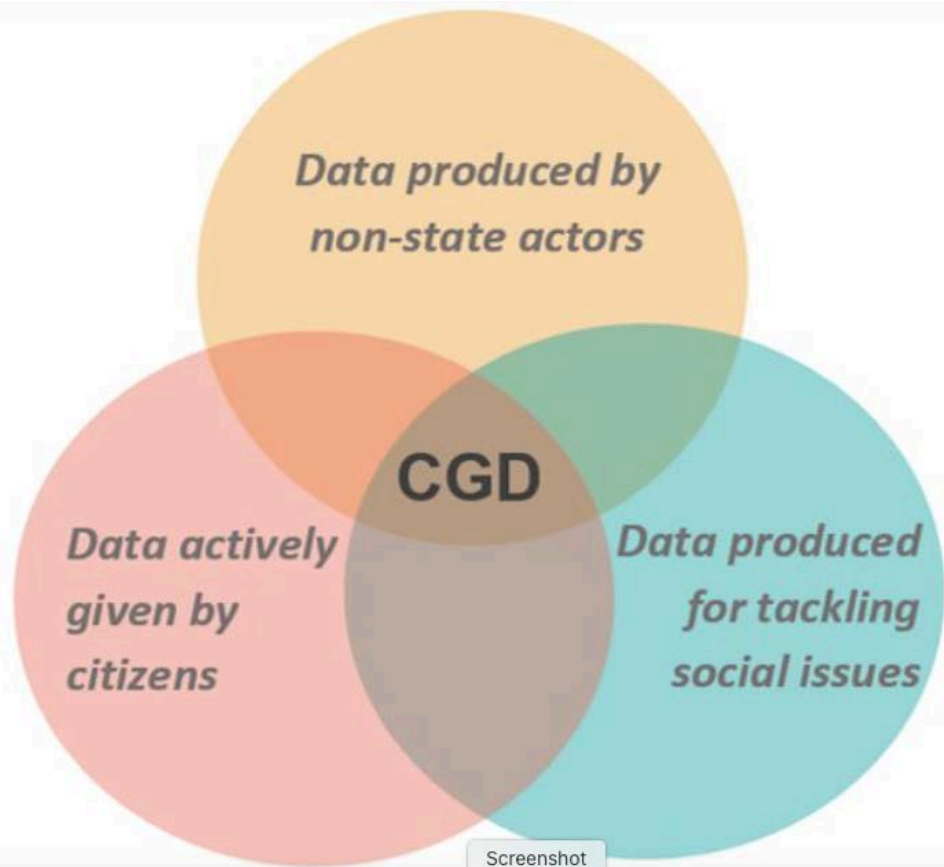
- GPSDD. Choosing and engaging with citizen generated data. https://www.data4sdgs.org/sites/default/files/services_files/Choosing%20and%20Engaging%20with%20CGD_The%20Guide_0.pdf

- Country explorations with PARIS21:

- Philippines. https://paris21.org/news-center/news/new-report-shares-insights-using-citizen-generated-data-sdg-reporting-philippines

- Maldives. https://paris21.org/news-center/news/citizen-generated-gender-data-maldives-connecting-data-ecosystems

---

[60] https://www.data4sdgs.org/resources/choosing-and-engaging-citizen-generated-data-guide

Systematic assessment and approach to capacity development: Integrating gender in National Statistical Development Strategies

The NSDS is the key instrument for communicating visions for improving data and statistics and systematically mobilizing resources. As part of its collaboration with UN Women, PARIS21 developed a comprehensive framework to assess data and capacity gaps linked to gender statistics. The framework, aimed at national statistical offices, proposes methods, activities and tools for conducting assessments related to gender statistics to support the mainstreaming of gender statistics in national statistical systems (NSS). The assessment report resulting from the application of this framework is meant to inform national strategies for the development of statistics (NSDS).

- PARIS21 and UN Women. 2020. Assessing data and statistical capacity gaps for better gender statistics: Framework and Implementing Guidelines. https://paris21.org/sites/default/files/inline-files/Framework%202020_update_web_0.pdf

- Coming soon: PARIS21 and UN Women. Guidelines for integrating gender in national strategies for sustainable development.

Box 6-3. Making gender statistics a top priority in the Senegalese NSDS

"Following an 8-months-long process, Senegal adopted its third national strategy for development of statistics (NSDS III) for 2019-2023 with the support of the African Development Bank and PARIS21. One of the most remarkable features of this strategy is the plan to expand the production and dissemination of gender statistics at the national level. This goal constitutes the first strategic pillar of the NSDS III, and aims to make official gender statistics more responsive to increasing demand." *Source: https://paris21.org/news-center/news/making-gender-statistics-top-priority-senegalese-national-strategy-development*

For more information: https://paris21.org/sites/default/files/inline-files/Framework%202020_update_web_0.pdf

Systematic assessment and approach to capacity development: CD 4.0

PARIS21 has launched a pioneering approach to capacity development — Capacity Development 4.0 – that goes beyond technical skills and places an emphasis on leadership, change management, advocacy and networking. Figure 6-4 is a schematic diagram of the framework.

These are the first guidelines to advise national statistical offices and development cooperation agencies on how to engage in country-led, sustainable and participative statistical capacity development. The guidelines explain the CD4.0 approach and present 30 activities on how to implement capacity development programmes, illustrated by case studies to provide real-world context.

Figure 6-4: Framework of Capacity Development 4.0

Target/Level	Individual	Organisational	System
Resources			
	• Professional background	• Human resources • Budget • Infrastructure	• Legislation, principles and institutional setting • Funds infrastructure • Plans (NSDS, sectoral…) • Existing data
Skills and knowledge			
	• Technical skills • Work know-how • Problem solving and creative thinking	• Statistical production processes • Quality assurance and codes of conduct • Innovation • Communication	• Data literacy • Knowledge sharing
Management			
	• Time management and prioritisation • Leadership	• Strategic planning and monitoring and evaluation • Organisational design • HR management • Change management • Fundraising strategies	• NSS co-ordination mechanisms • Data ecosystem co-ordination • Advocacy strategy
Politics and power			
	• Teamwork and collaboration • Communication and negotiation skills • Strategic networking	• Transparency • Workplace politics	• Relationship between producers • Relationship with users • Relationship with political authorities • Relationship with data providers • Accountability
Incentives			
	• Career expectations • Income and social status • Work ethic and self-motivation	• Compensation and benefits • Organisational culture • Reputation	• Stakeholder interests • Political support • Legitimacy

For more information: PARIS21. 2020. Guidelines for Developing Statistical Capacity- A Roadmap for Capacity Development 4.0. https://paris21.org/sites/default/files/inline-files/UNV003_Guidelines%20for%20Capacity%20Development%20PRINT_0.pdf

Challenges of data disaggregation and options for improvement—assessment, use of ADAPT, examining options for improvement and integrating into NSDS— Country experiences

- Cambodia @ https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%202.b.1_Cambodia_Experience%20in%20producing%20disaggregated%20data%20for%20SDGs.pdf

- Ethiopia @ https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%202.b.2_%20Ethiopia%20Data%20Disaggregation%20Practice%20on%20National%20SDG.pdf

- Turkmenistan @ https://unstats.un.org/sdgs/files/meetings/sdg-inter-workshop-jan-2019/Session%202.b.3_Turkmenistan_Bang_270119%20ENG%20PRINT.pdf

What's Next

This Guidebook has two versions—an initial printed publication and an online version that will be a "living document" constantly updated to keep up with the growing work and experiences that generate useful knowledge for producing and analyzing data to leave no one behind. In the very near future, for example, readers can expect two significant compilations of tools and guidance materials:

- Based on an IAEG-SDG/UNSD stock-taking questionnaire to established city and expert groups under the United Nations Statistical Commission, international expert groups/committees, as well as UN agencies and stakeholders focusing on vulnerable populations, disaggregation dimensions and data collection methods.

  @ https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-11/Compilation%20of%20tools.guidance%20of%20existing%20materials%20for%20data%20disaggregation-%20DRAFT.pdf

- Coming in March 2021. UN Women and ISWGHS. 2021. Counted and Visible: Toolkit to Better Utilize Existing Household Surveys to Generate Disaggregated Gender Statistics.

  The Counted and Visible Toolkit is a collection of good practices as well as learnings of select country studies, each focusing on specific aspects of the statistical process aimed at ensuring a holistic, sustainable, and institutionalized approach of producing disaggregated gender statistics using existing data from household surveys. The stages were also guided by the overarching aims of UN Women's global gender data programme, Women Count, of ensuring an enabling environment, increasing data production, and increasing access and use to inform policies.