Regional Forum on Education Policy / vol. 5 / 2021

Use of Education Management Information Systems (EMIS) for effective monitoring SDG4 at national, regional, and global level

Shailendra Sigdel
Statistical Advisor for South Asia
UNESCO UIS

About this publication

This document, part of the presentation carried out in the Regional Forum on Education Policy in 2021, summarizes and expands on Use of Education Management Information Systems (EMIS) for effective monitoring of SDG4 at national, regional, and global level. This report elaborates on the processes followed by the UNESCO Institute for Statistics (UIS) as the custodian agency of SDG 4 to aggregate and curate the data. To underpin the findings from the presentation, additional context, definitions, and data is included in this report, which is primarily drawn from three reports viz., Operational Guide to Using EMIS to Monitor SDG 4 (UNESCO-UIS, 2020a), Data Innovation for Producing SDG 4 Indicators: A Global Analytical Report (UNESCO-UIS, 2020b) and Buyer's and User's Guide (UNESCO-UIS & GPE, 2020).

How to cite this publication: Shailendra, S. (2022). Use of Education Management Information Systems (EMIS) for effective monitoring SDG4 at national, regional, and global level. IIEP UNESCO.

Index

1. Background

Increasing the availability and use of data and evidence is the critical arena for effective management of the education sector. Low-and middle-income countries face challenges in improving their education systems' ability to meet the ambitious goals of Sustainable Development Goals 4 (SDG 4) related to access, quality, and equity in education. Thematic indicators based on the Education Management Information Systems (EMIS) helps countries monitor progress on SDG 4 indicators. EMIS facilitates the collection, processing, analysis and dissemination of education data and information to support the monitoring and performance evaluation of an education

EMIS

system. Education systems are multifaceted and complex and EMIS efforts must be pulled up, to produce quality, disaggregated and timely data for informed policy decisions (UNESCO-UIS, 2020b). The COVID-19 pandemic has led to unprecedented levels of disruption in education due to nationwide and localized closures, that have impacted hundreds of millions of students. Especially, during such crisis, when the most vulnerable are disproportionately impacted, it is imperative to make right use of data to introduce tailored programs and policies to ensure continuation of learning for students.

The Incheon Declaration and the SDG Education 2030 Framework for Action recognizes the importance of good quality disaggregated data, broken down by specific characteristics of given population groups to ensure that indicators can monitor progress towards reducing inequality (UNESCO, 2015). Systematic data on education at different levels are scarce, and often their quality cannot be attested. The production of data on programs and enrolment, learning outcomes, as well as financing and issues related to the identification of the target populations remains a key challenge in the education sector. Ultimately, education systems only function effectively if their strategies, approaches, and funding are built on a solid foundation of quality data managed in efficient information systems (UNESCO-UIS, 2020a). A methodical understanding and use of EMIS plays a pivotal role in producing comparable data for effective monitoring of SDG4 at national, regional, and global level. Thus, this report sheds light on the global status of EMIS and how the EMIS in the LAC region fares in compared to other regions and delves into some of the specifics of data collection through EMIS. This report further analyses the pertinent challenges in disaggregation of education data and proposes a way forward to mitigate these challenges for effective uptake of EMIS at all levels.

2. Objective

This report highlights the key characteristics of existing national EMIS to illustrate the type of EMIS, level of data collection and capacity of EMIS to produce administrative data related SDG4 global, regional and thematic indicators. The document further proposes a way forward to support countries in their efforts to collect and produce standardized quality data. The primary objective of this summarized presentation report is to assist countries in Latin America develop capacity to understand the process of data collection for EMIS, availability of education data across regions, increase familiarity with the indicators and variables used to extract data from EMIS system, and understand existing challenges during data disaggregation. It is expected that the elucidations and recommendations in this report will assist countries in developing an integrated EMIS and in navigating a way forward for effective use of EMIS to monitor progress of their respective country towards SDG 4.

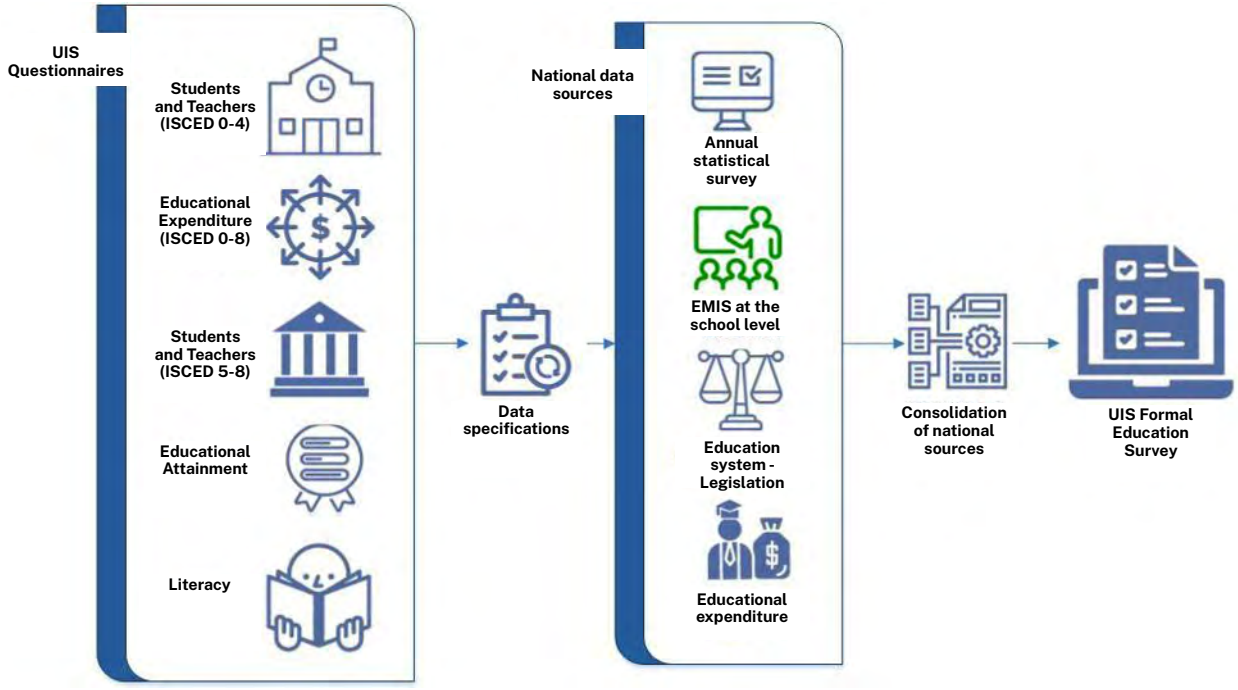
3. Data Collection Methodology

3.1. Process of EMIS Data Reporting to UIS

National education systems differ across countries and region in terms of both structure and curricular content. To facilitate comparison of data between countries, UNESCO/UIS has developed the International Standard Classification of Education (ISCED 2011) for cross-country comparisons of education statistics. The ISCED 2011 classification helps benchmark performance across countries over time and monitor progress towards national and international goals (UNESCO-UIS, 2012). Therefore, as a first step, to ensure that data on enrolment in primary education and other indicators are comparable, national data must be mapped to the ISCED. UIS collects data from students and teacher (ISCED 0-4, Questionnaire A), education expenditure (ISCED 0-8, Questionnaire B) and student and teachers (ISCED 5-8, questionnaire C), education attainments and literacy by using UIS standard questionnaires. These data come from National data sources that

includes the annual statistical survey, EMIS at the school level, education system-legislation, education expenditure data etc. Once the data from national data sources is consolidated it feeds into the UIS Formal Education Survey. Mapping datasets from different countries to the same ISCED classification helps to understand and properly interpret the inputs, processes, and outcomes of education systems from a global perspective (UNESCO-UIS, 2012).

Figure 1. Process of EMIS data reporting to UNESCO Institute for Statistics (UIS)



The member states administer annual statistical surveys to gather data on different aggregates for statistical purposes. Once member states fill out the UIS survey forms, the forms are received by UIS, and the quality of data is scrutinized for quality assurance. The data from these surveys are examined comprehensively through a reiterative review process and consultation between the UIS and the member state before final publication. The data pertaining to household surveys (education attainment and literacy) are produced by national statistical agencies and are not normally included in a traditional EMIS. Similarly, the education expenditure data are derived from the budget book of the ministries of finance and other sources (UNESCO-UIS, 2020a). Figure 1 illustrates the process currently in place for UIS Questionnaires and Consolidation of National Sources.

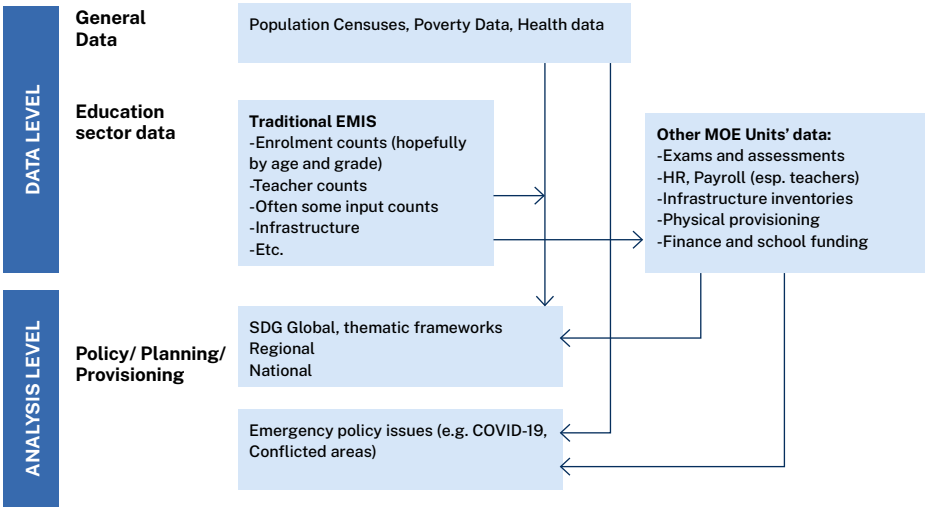
For quality control of data, the UIS continually works with countries to harmonize the data through multiple rounds of review and finalization of the indicators. Being the custodian agency for SDG4 monitoring, UIS also provides support to the member states on streamlining methodologies and building capacity of member states as well.

3.2. Sources of Data

To produce internationally comparable education indicators, the UIS uses several main sources of data, and these sources differ in coverage, timing, and the way they are collected. There are four primary data sources used in calculating the Sustainable Development Goal (SDG) 4 indicators. These include household-based survey and assessment data, census data, school-based survey and assessment data, and Education Management Information System (EMIS) or administrative data. The school/individual based administrative data can produce around half of the 43 thematic indicators (UNESCO-UIS, 2020b).

At the preliminary data level, general data is retrieved from population censuses, surveys and through health and poverty data. More specific education sector data such as enrollment counts, teacher counts, school infrastructure are based on information collected in the management of the education system. Within the education system, data are collected from the three units i.e., school, student, and teacher – and stored on a data management platform. These are usually available from EMIS and used by Ministries of Education for management and planning purposes and are typically updated on an annual basis (UNESCO-UIS, 2020a). Some administrative data, such as exams and assessments, human resource payroll (teachers' salaries), infrastructure inventories, physical provision, finance, and school funding can be retrieved from other from non-EMIS sources in Education Ministries (see Figure 2). After data is collected, at the analysis level, when it comes to policy, planning and provisioning, SDG Global and thematic indicator frameworks, regional and global policies are developed. Latin America is currently moving forward to develop a regional monitoring framework. These indicator frameworks allow countries to map their progress according to their own national plan but also allows for comparison with the global and thematic indicators. In the aftermath of the COVID-19 pandemic and emergency policies introduced, school attendance was affected globally as many children were out-of-school. Emergencies as such poses challenges on managing and provisioning for the nature of data to be collected through EMIS.

Figure 2. EMIS: main source of data for monitoring SDG4



EMIS

4. Global Context on EMIS data collection

4.1. Response Rate to EMIS Typology Questionnaire

To gain a better understanding of the characteristics of the existing EMIS database globally, a questionnaire was developed, translated in multiple languages, and subsequently disseminated to UIS counterparts across all regions. Based on this EMIS metadata survey conducted in 2020, data was accumulated from 103 countries out of the 175 countries in the region. The questionnaire was sent to UIS counterparts covering sub-Saharan Africa, the Arab States, South and West Asia, Central Asia, East Asia, the Pacific, Latin America and Caribbean, and Central and Eastern Europe regions. As seen in figure 3, the overall response rate of the questionnaire was 59% across all regions. The response rate from the Latin America and Caribbean (LAC) region stood higher than average where 65% where 24 out of 37 countries in the region responded.

Figure 3. Response rates of EMIS Typology Questionnaire by Region

Regiões	Nº de países na região	Nº de países que responderam o questionário	% de países (respostas ao questionário)
América do Sul	12	9	64
América Central	7	6	86
Caribe	16	9	56
América Latina e Caribe	37	24	65
Média mundial	175	103	59

Source: UIS (2020). EMIS Metadada Survey.

4.2. Region-wise Data Management Platform Used

A data management platform is used to collect, process, and analyse data from three major components of the education system i.e., the school, teacher, and student, to inform SDG 4 indicators. It is crucial for tracking changes, ensuring data quality and timely reporting of essential information for planning and management, and for facilitating the use of information in decision making by policymakers. Data management platform varies across regions based on country context such as availability of resources; location of decision making (province/state level); requirements of disaggregated data and culture of using of evidence in planning and monitoring etc (UNESCO-UIS, 2020a).
It is interesting to note that the primary data management platform used by majority of the countries is their own developed data management software. On average, 80% of countries developed their own data management software and 18%, 13% and 5% of countries use the StatEduc Software, OpenEMIS and other tools such as Excel respectively. In Latin America and Caribbean, 83% of the countries reported using their homegrown software whereas 21% use built-in software like OpenEMIS for data management. It is important to note that some countries use multiple types of software within the MOE or between different departments. Since the use of data management platform is not exclusive to use of one software, the percentage of countries using different software does not add up to 100 percent in the adjacent Figure 4.

Figure 4. Data Management Platform by Region

Regions	Own developed	StatEduc	Ed Assist	OpenEMIS	Others (e,g. Excel)
South America	89	0	0	11	0
Central America	83	17	0	17	0
Caribbean	78	11	0	33	11
Latin America and Caribbean	83	8	0	21	8
World Average	80	18	0	13	5

Source: UIS (2020). EMIS Metadada Survey.

4.3. Mode of Data Collection

Paper, online interface, and other standalone electronic modes are the most prevalent modes of data collection. Figure 5 illustrates the breakdown of the mode of data collection adopted by each region. It can be observed that paper is the predominant mode of data collection. The data shows that 53% of the countries across all regions primarily use paper for data collection, 51% of countries use online interface and 36% of countries use standalone electronic mode (e.g.: Excel and PDF formats). Some countries use multiple modes of data collection depending on the level of education or unit of data collection (e.g., school, teacher, and student) as well (UNESCO-UIS, 2020b). Technological advancement and spread of internet have had a huge implication on EMIS. Substantial number of countries have switched from using solely paper for data collection purposes to other online or offline digital means. While internet access is available in all large cities, major towns and urban schools in most countries, most rural areas and rural schools in several countries have limited or no internet access. The accessibility of internet in schools varies across regions. Therefore, due to the lack of availability of internet facilities and other resources many countries still opt. to use paper to collect data from rural schools. The data gathered through offline means is entered into the computer system later, either at the national level or at district headquarters, depending on wherever internet and human resources are available (UNESCO-UIS, 2020b). It is important to note that with the accelerating expansion of internet and technological tools, the uptake of online and digital mode of data collection is bound to increase which can positively impact the ease and efficiency of data collection.

Figure 5. Mode of Data Collection (% of countries)

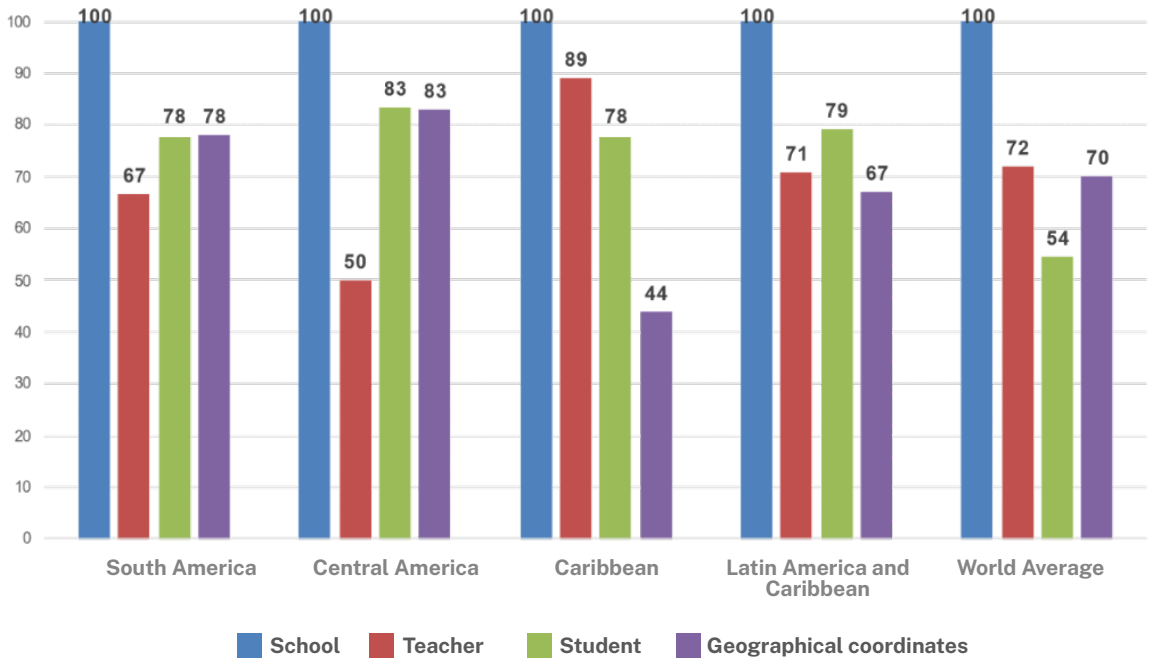
Regions	Paper (% of countries)	Standalone electronic mode (% of coutries)	Online interface (% of countries)
South America	33	33	89
Central America	33	17	67
Caribbean	56	56	56
Latin America and Caribbean	42	38	71
World Average	53	36	51

Source: UIS (2020). EMIS Metadada Survey.

4.4. Data Recording in Each Region

The first unit of data recording is done at the school-level in a country's EMIS database. Figure 6 shows that 100% of countries record school-level data electronically in the national EMIS database. It is important to collect administrative school-level data as better school conditions are helpful for delivering a quality education, which in turn helps education policymakers take evidenced-based policy decisions and inform international reporting on SDG 4 indicators (UNESCO-UIS, 2020b).

Figure 6. School, teacher, student and coordinates level data recording in EMIS (% of countries)



Source: UIS (2020). EMIS Metadata Survey.

The second unit of data recording is at the student-level in the EMIS database. It is essential to record student-level data as it includes information such as enrolment, new entrants, attendance, transfers, and dropouts. This data provide insight into student progression which policymakers and other stakeholders use to take policy decisions. Based on student-level data, countries can not only identify gaps and prioritize resources, but also inform international reporting on SDG 4 indicators. Figure 6 depicts that there is a significant drop in student level data from school level data (UNESCO-UIS, 2020b). On average only 54 percent of countries record individual student-level. However, it is interesting to note than percent of countries recording individual student-level in the LAC region is at 79%, which significantly higher than global average.

The third unit of data collection in the EMIS database is teacher-level data. By 2030, the SDG 4.c target aims to increase the supply of qualified teachers through international cooperation for teacher training in developing countries and small island developing states. Hence, it is essential to not only strengthen the capacity of teachers in each country but also to monitor and track the progress of training, the pupil-teacher ratio and other parameters related to teachers. Individual

teacher-level data helps to increase data quality and relevancy and allows reliable disaggregated analysis for policy interventions (UNESCO-UIS, 2020b). Figure 6 represents that on average 72% of countries record individual teacher-level data and similarly 71% of countries in the LAC region recorded teacher level data.

Higher percentage of countries record school-level data compared with individual student or teacher-level data as most countries have been recording school-level data for decades and already had a system in place. However, many countries have either recently started collecting data at the individual student and teacher levels or are still in the process of adopting mechanisms to collect data at the individual student and teacher levels (UNESCO-UIS, 2020b).

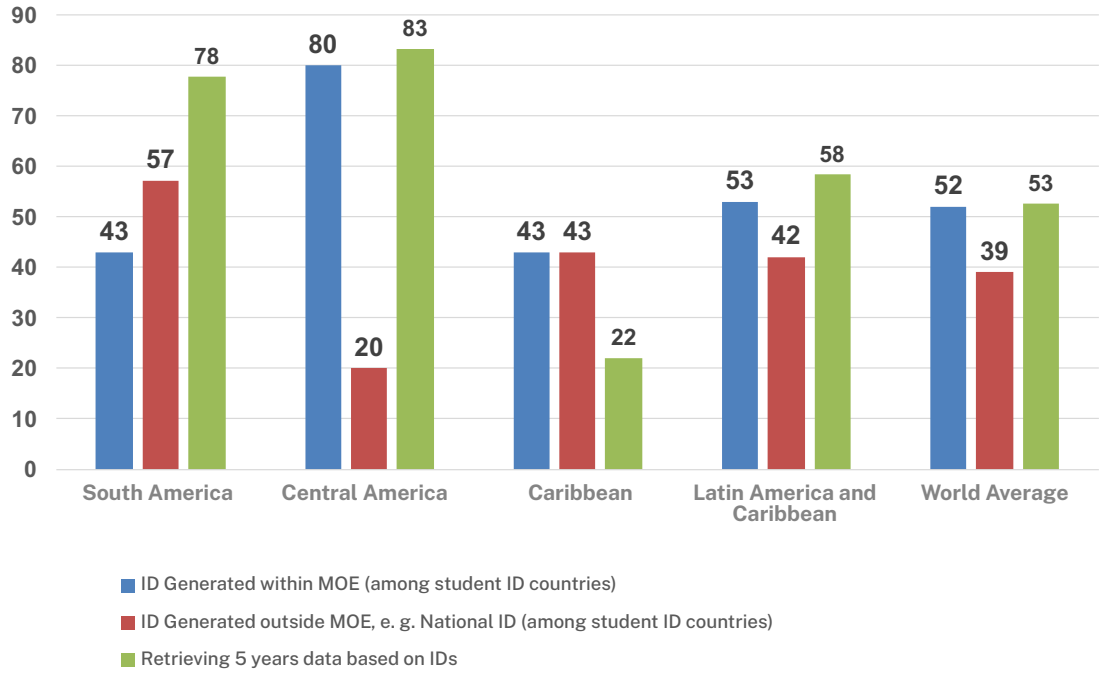
Finally, a novel unit of data collection within the EMIS database is geographical coordinates. The geographical coordinates of a school when linked with other spatial data is useful for targeted educational planning, management, and monitoring. The geographical coordinates build a geospatial database with a relational database of educational, demographic, social and economic information. The coordinates can provide access to multiple sources of essential data such as census data, and other ancillary data from transportation, utilities, healthcare, household incomes, employment, land use, and agriculture. Therefore, collecting geographical coordinates of schools is important as it helps policymakers take decisions based on a holistic picture of different variables (UNESCO-UIS, 2020b). Out of 100% of countries that record school level data, only 71% of them are collecting geographical coordinates of schools. In the LAC region 67 % of the countries report geographical coordinates. The reason for the low coverage is because countries have either only recently started collecting geographical coordinates of schools and hence, have not been able to cover all schools, or plans are underway to integrate the geographical coordinates in its EMIS database (UNESCO-UIS, 2020b).

4.5. Student ID generated in each region

Primarily, there are two types of student IDs used by different countries. First, a student ID generated by the EMIS system by the MOE and second, an ID generated outside the MOE, for example, a national ID. In most cases, ID generated under the MOE is used only within the MOE for data collection and other management purposes such as the transfer of scholarships, preparation of certificates of examination results etc. Student ID generated outside the MOE, for example a national ID, is used outside the MOE as well as within the MOE database. Its compatibility is much higher than MOE generated ID since the national ID can be linked with many other government and non-government databases outside of MOE, such as social benefits, direct cash transfer, family incomes, health data etc. It helps provide an intersectoral analysis for the development of intersectoral policies (UNESCO-UIS, 2020b). For instance, the COVID-19 pandemic has had an enormous impact on both the health and education sector and the impediments on one sector has undoubtedly affected the other. It is evident that data from these sectors cannot be analysed in silos. If data from Health Management Information System (HMIS) and EMIS can be interlinked through a unique ID generated outside of MoU, such as a national ID, it could yield much more apt and targeted program interventions.

Figure 7 illustrates that globally on average, only 52% of the countries use a unique student ID generated within the MOE and 39% generate a unique student ID outside the MOE. This proportion varies across regions. In the LAC region, 53% of the countries responded that they use an MOE generated student ID and 42% reported that they have a unique student ID generated outside of the MOE, which is slightly higher than the world average.

Figure 7. Generation of Student ID by Region (% of countries)



Source: UIS (2020). EMIS Metadata Survey.

In terms of retrieval of past student data, Figure 7 shows that, on average, only 53% of countries can retrieve five years of student data based on student ID. This percentage is slightly higher when it comes to countries in the LAC region. It is reported that 58% of countries in the LAC region can retrieve 5 years of student data based on IDs; Retrieval of 5 years of student data is vital to track the progress of a student or a cohort of students. It helps to make informed policy decisions, prioritize resources, and monitor progress. It also indicates the sustainability, integration, and functionality of student-level data in the EMIS. There are two reasons for low data retrieval. First, most countries started collecting data at the student level in the last five to six years. Second, countries use multiple types of software, and system incompatibility leads to difficulties in system migration and integration of data (UNESCO-UIS, 2020b).

5. EMIS Variables, SDG 4 Indicators and Unit of Data Collection

5.1. Mapping of EMIS Variables to SDG4 Indicators

Creation of global and thematic indicators were designed with the objective of identifying key concepts that need to be measured to monitor progress towards achievement of SDG4 and whether existing policies would be sufficient to ensure that the target could be met or whether further action, including remedial action is needed to get back on track. Traditionally, EMIS are prepared to report on indicators referred to as administrative data. Many of the indicators reported on using other sources, such as learning assessments and household surveys, are not usually embedded in a traditional EMIS (UNESCO-UIS, 2020a). The variables in Figures 8, 9 and 10 represent the administrative data that needs to be collected through an EMIS from schools for the estimation of SDG4 Indicators. Figure 8 represents the mapping of variables from EMIS to SDG 4 and identifies which

variable is linked to more than one SDG 4 indicator. For instance, as seen in Figure 8. the Enrolment variable from EMIS is linked with 9 SDG4 indicators (SDG4: 4.1.4, 4.1.5, 4.2.2, 4.2.4, 4.3.2,4.5.2,4.5.4, 4.c.2 and 4.c.4). All these variables are required to monitor SDG4 effectively. It is important to note that the variable mapping of the indicators does not include system-level indicators, such as compulsory years of schooling nor indicators that require school-or student-based surveys and assessments such as attacks on school and minimum proficiency levels etc. (UNESCO-UIS, 2020a).

Figure 8. Variables captured by EMIS under SDG 4

Variable	SDG 4 Indicators
Enrolment	4.1.4, 4.1.5, 4.2.2, 4.2.4, 4.3.2, 4.5.2, 4.5.4, 4.c.2, 4.c.4
Graduates from an education level	4.1.2
New entrant students to a education level	4.1.2
New entrant students to the last grade of each education level	4.1.3
Participation of youth and adults in format and non-formal education and training	4.3.1
Participation of youth in technical and vocational education and training	4.3.3
Participation of youth and adults in literacy programmes	4.6.3
Schools	4.7.2, 4.a.1
Schools providing life skills-based HIV and sexuality education	4 .7.2
Schools with and without access to given facilities	4.a.1
Teachers	4.c.1, 4.c.3, 4.c.6, 4.c.7
Qualified teachers	4.c.3, 4.c.4
Teachers with the minimun required qualifications	4.c.1
Trained teachers	4.c.2
Teachers who recieved in-service training of each type in the last year	4.c.7
New entrant teachers	4.c.6

Source: UIS (2020). Operation manual.

5.2. Variables Required for Data Disaggregation and Unit of Data Collection

Depending on the data required to inform the SDG4 indicators, the variable data needs to be collected from one or more unit of data collection. For instance, to collect data for SDG 4.1.2.-Completion Rate, the data needs to be collected from either individual student level or school level data on graduation and new entrants at primary, lower secondary and upper secondary level (Refer to Table in Figure 9). In terms of disaggregation, the minimum disaggregation required for SDG 4.1.2 is by sex.

Figure 9. Variables required for Data Disaggregation and SDG4 Indicators (Table 1)

SDG 4 Indicators	Unit of data collection	Variable	Formal education								Disaggregation				Possibility by level of education	Population
			ISCED Level								Essential		Important			
			01 Early childhood	02 Pre-primary	1 Primary	2 Lower Secondary	3 Upper Secondary	4 Post Secondary not tertiary	5 Short cycle tertiary	6-8 Bachelor to PhD	Age/age-range	Grade	Sex	Type of institution	Level of education	
4.1.2. Completion Rate	Individual student / Level	Graduates from an education level			x	x	x						x		Yes	No
	Individual student / Level	New entrant students to a education level			x	x	x						x			
4.1.3 GIR to last grade	Individual student / Grade	New entrant students to the last grade of each education level			x	x						x	x		No	Yes
4.1.4. Out-of-school	Individual student / Level / Age	Enrolment			x	x	x				x		x		No	Yes
4.1.5. Children over -age	Individual student / Grade / Age	Enrolment			x	x					x	x	x		Yes	No
4.2.2. Particip. before Primary	Individual student / Level / Age	Enrolment	x	x	x						x		x		No	Yes
4.2.4. GER Early Childhood	Individual student / Level	Enrolment	x	x									x		No	Yes
4.3.1. Youth/adults educ./training (formal / Non-formal)	Individual student / Level / Age	Participation of youth and adults in format and non-formal education and training									x		x		No	Yes
4.3.2. GER Tertiary	Individual student / Level	Enrolment							x	x			x		No	Yes
4.3.3. Youth/adults TV educ/training	Individual student / Orientation / Age	Participation of youth in technical and vocational education and training				x	x	x	x		x		x		No	Yes
4.5.2. 1st/home language	Individual student / Level	Enrolment			x								x		Yes	No
	Individual student / Level	Enrolment by first or home language			x								x			
4.5.4. Expediture per student	Individual student / Level	Enrolment	x	x	x	x	x	x	x	x			x	x	No	No
4.6.3. Youth/adults in literacy prog.	Individual student / Level / Age	Participation of youth and adults in literacy programmes									x		x		Yes	Yes

Figura 10. Variables required for Data Disaggregation and SDG4 Indicators (Table 1 contd.)

4.7.2. Schools x/ sexuality education	Individual schools / Level	Schools			x	x	x								Yes	No
	Individual schools / Level	Schools providing life skills-based HIV and sexuality education			x	x	x									
4.a.1 Schools w/ basic services	Individual schools / Level	Schools			x	x	x								Yes	No
	Individual schools / Level	Schools with and without access to given facilities			x	x	x									
4.c.1 Teachers w/ min. qualifications	Individual teachers / Level	Teachers		x	x	x	x	x					x		Yes	No
	Individual teachers / Level	Teachers with the minimun required qualifications		x	x	x	x	x					x			
4.c.2 Pupil-trained teacher	Individual teachers / Level	Trained teachers		x	x	x	x	x					x		Yes	No
	Individual student / Level	Enrolment		x	x	x	x	x					x			
4.c.3 Teachers qualified	Individual teachers / Level	Teachers		x	x	x	x	x					x		Yes	No
	Individual teachers / Level	Qualified teachers		x	x	x	x	x					x			
4.c.4 Pupil -qualified teacher	Individual student / Level	Enrolment		x	x	x	x	x					x		Yes	No
	Individual teachers / Level	Qualified teachers		x	x	x	x	x					x			
4.c.6 Teacher attrition rate	Individual teachers / Level	Teachers		x	x	x	x	x					x		Yes	No
	Individual teachers / Level	New entrant teachers		x	x	x	x	x					x			
4.c.7 Teachers w/in -service training	Individual teachers / Level	Teachers		x	x	x	x	x					x		Yes	No
	Individual teachers / Level	Teachers who recieved in-service training of each type in the last year		x	x	x	x	x					x			

In similar manner, for population related indicators, such as 4.1.4.-Out of School children, both the individual student level data and population data is required. In terms of disaggregation, for the 4.1.4 indicator, disaggregation is required by both age and sex. For this indicator, data must be collected at the individual student level and cannot be collected just at the level of school education. It is important to note that not all indicators can be reported at the school level (e.g., location) and they cannot be disaggregated. It is necessary to collect specific variable(s) from schools to calculate indicators and some of the indicators based on administrative data require age-specific population data other than those collected from schools (UNESCO-UIS, 2020a). The different types of disaggregation by sex, age, types of disability, location, and ethnicity is helpful when the indicators are ultimately used to inform targeted policy interventions.

EMIS

6. Challenges in Data collection, Disaggregation, and Variables

As stated in the Preamble of UN resolution 70/1 of September 2015, 'Leaving no one behind' is a fundamental tenet of the 2030 Agenda for Sustainable Development, and the Sustainable Development Goals. The monitoring of the SDG agenda increases the demands on EMIS for disaggregated data. SDG Target 4.5, for example states: By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples, and children in vulnerable situations (UN General Assembly, 2015). To ensure that no one is left behind while attaining development goals, it is imperative to gather disaggregated and quality data. Aggregated data or data summaries often hide the discrepancies in access to education that exist between disparate groups. For instance, schooling in an urban region might differ from schooling in a rural area. Likewise, school enrolment and attendance of girls versus boys may differ in the same school. On the other hand, disaggregated data highlights the distinct characteristics of diverse groups which allows policies and interventions to effectively provide support to marginalized and vulnerable populations. Much information currently produced through national EMIS would be more useful if recorded in a harmonized way to facilitate classification, comparability and use over time. Globally, all countries collect data from schools on an annual basis, and many developing countries have recently started collecting information on individual data on teachers and students, which provides an opportunity to disaggregate the indicators (UNESCO-UIS, 2020b). However, when it comes to data disaggregation, the process and level of disaggregation is not always straightforward. There are multiple challenges to understand what data can be and should be disaggregated. Some of the common challenges in disaggregation are specified and elaborated below.

Figure 11. Key Challenges in Disaggregation

Disaggregation and challenges
1. **Sex:** female / male / third gender
2. **Age:** based on official document. Viral registration is not universal, based HT's judgement
3. **Location:** definition of 'urban/rural', comparability, keeps changing
4. **Disability:** Standard definition, Acid affected
5. **Ethnicity:** do not want disclose or want to be counted
6. **Wealth:** Data quality
7. **Migration:** forced migration (displaced), internal migration
8. **Affected:** Conflict affected, disaster affected
9. **Language:** home language / multiple languages

1. Disaggregation by Sex. The sex of a person denotes the biological difference between male and female, which is typically assigned at birth. Data disaggregation by sex is essential for education-related gender statistics as it provides the context for most analyses of the social and demographic characteristics of a population (UNESCO-UIS, 2020a). Globally, most EMIS collect disaggregated data by male and female but some countries like Canada and Nepal collect third gender data. However, disaggregating by sex which is a binary concept is outmoded and is not inclusive for people who choose to identify as a different gender than the one which is assigned at birth. Disaggregating data into previously undefined is necessary since the collection and analysis of gender data across the SDG4 indicators is critical for gauging progress on global gender equality in terms of access to quality education.
2. Disaggregation by Age. The age of a person is the interval of time between their date of birth and the reference date of the data collection, expressed in completed solar years. For precision, recording the date of birth (year, month, and day) helps establish age accurately, especially when EMIS data collection occurs more than once during the same school or academic year. Data disaggregation by single age (and grade) is essential for education-related age statistics as it provides the context for most analyses of the social and demographic characteristics of the population (UNESCO-UIS, 2020a). Disaggregating data by age is challenging as age groupings used to record and report education data varies, which hampers the comparability, and efficacy of these data. A standardized universal approach in recording age data is not followed throughout all regions and countries. Age can be based on either an official document, vital registration, or based merely on an ad-hoc judgement of the head teacher at a school. Vital registration is not uniform across countries. Especially in low-income and middle-income countries, many people have never had their births registered (UNICEF,2019). In absence of official documents reporting age, the data may not be accurate and reliable.
3. Disaggregation by Location. It is apparent that certain regions and locations such as rural areas have less accessibility to quality education in compared to urban areas. Currently, there is no international standard definition for 'urban' and 'rural.' Therefore, records in EMIS of information by location is aligned with national definitions. National definitions of 'urban' and 'rural' are usually established by the relevant government agency, such as the ministry of interior, urbanization, territorial administration, or infrastructure (UNESCO-UIS, 2020a). However, disaggregated data by location is often not harmonized and comparable even within the same country. Based on the criteria set by the delegated government agency of a particular country, a location can be demarcated as rural in one year, and the same location can be established as an urban area in the following year. Especially in countries where political leadership is volatile, the demarcation of rural and urban areas keeps changing, which affects collection of uniform and comparable data.
4. Disaggregation by Disability Status. Student with disabilities can be defined as 'persons who are at greater risk than the general population for experiencing restrictions in performing specific tasks or participating in role activities' (UNSD 2017). Underrepresentation in schools throughout the world is particularly pronounced among children with disabilities. Children with disabilities globally make up at least one-third of out-of-school children (UNESCO, 2016). Having reliable, accurate, and comparable data that can be disaggregated by learners' disability status is essential to effectively monitor progress in reaching marginalized children and to provide the information needed to continue to strengthen the quality of education for these learners (USAID, 2020). At present the data available in many EMIS are limited for students with disabilities in education institutions. Data on disability status should ideally be collected for

each individual student but in most cases, disability data is currently aggregated by grade at the school level (UNESCO-UIS, 2020a).

Similarly, the definition of disabilities is not uniform across countries. While many countries use standard definition of disabilities, others have unique definitions of disabilities. For example, India has identified 21 types of disabilities, where acid affected victims are also identified as persons with disabilities. This variance in understanding of disability status and lack of standard definition across countries makes it challenging for cross country data comparison and analysis.

5. Disaggregation by Ethnicity. Data on ethnicity provide information on the diversity of a population and can serve to identify subgroups of that population which can eventually help in providing targeted interventions to certain subgroups. Ethnicity can be defined as the state of belonging to a social group that has a common national or cultural tradition. Changes in ethnic identity can occur at both the group and individual level over a period. To harmonize management information system (MIS) data and record a student's ethnic origin or indigenous status, questions that capture certain criteria are proposed to facilitate consistent data collection across national educational institutions during each annual school census Depending on the country, questions on ethnicity and indigenous status could be answered through self-declaration or by providing option to choose from multiple ethnic affiliations. Where ethnicity represents a sensitive characteristic, appropriate data protection and disclosure control measures must be put in place (UNESCO-UIS, 2020a).

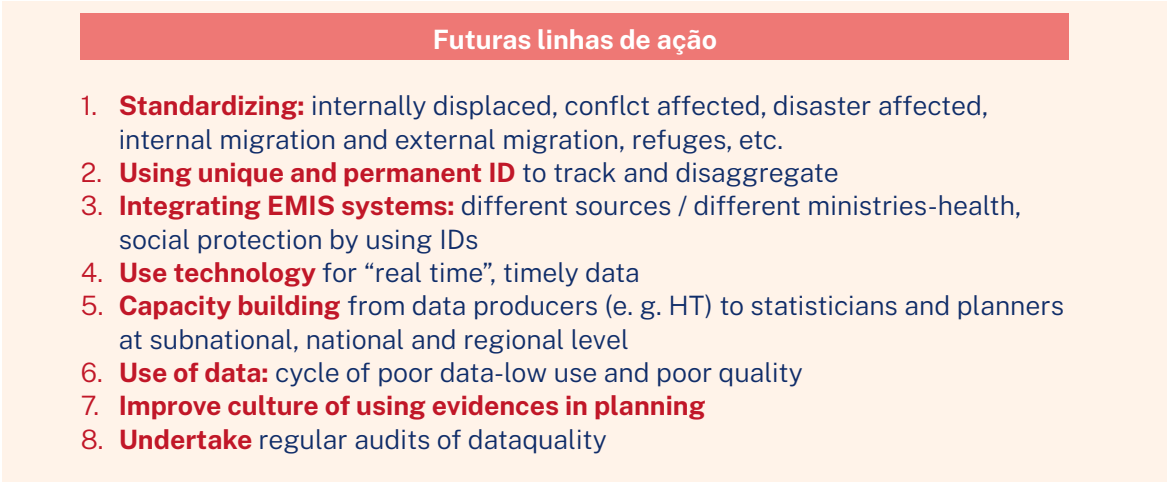
   However, the challenge with collecting and disaggregating data by ethnicity is that in many cases, individuals do not want to disclose their ethnicity due to fear of stigmatisation or facing certain ethnicity-based discrimination. While in some cases individuals want to be counted in multiple ethnicities to gain access to certain government allotted benefits and subsidies. This inconsistency in data collection creates challenge in controlling the authenticity of data.
6. Disaggregation by Wealth. Wealth is a characteristic typically related to adults or households. It is determined using two main models: the income model and, mostly in developing countries, the assets model. Information required from the respondent to determine wealth through these two models is both complex and sensitive. There is not enough information in general collected through the EMIS to accurately define the wealth status of a student. This is done in general using National Statistical Office information and household surveys (UNESCO-UIS, 2020a). Since data on wealth may require self-reporting of sensitive data from individuals, the quality and reliability of data on wealth cannot be attested. While some countries have developed certain criteria in gathering data on wealth, how to collect wealth data within EMIS it is still a widespread challenge.
7. Disaggregation by Migration. International migration and internal migration, both pose different challenges to education systems. The stock of international migrants present in a country is defined as 'the set of persons who have ever changed their country of usual residence, which is to say, persons who have spent at least a year of their lives in a country other than the one in which they live at the time the data are gathered' (UNSD, 1998). Apart from international migration and refugees, there is also a good number of students who migrate internally within a country for schooling. It is quite challenging for EMIS to record such migrant student; yet it is very relevant to policy and planning. Similarly, there persists the issue of inconsistent definitions where definition of forced migration and definition of internal migration may vary between countries. Thus, standardizing this definition is important for comparability. UIS is currently working to map definitions from different sources and in standardizing definitions to be used.

8. Disaggregation by Affected. Many of the largest education gaps are found in conflict-affected settings. Conflict-affected countries include over 20% of all children of primary school age, but account for around half of all out-of-school children of primary school age (UNESCO, 2011). The COVID-19 pandemic has further revealed the pressing need for accurate, disaggregated, reliable and timely data for Education in Emergencies (EiE). Despite the pandemic affecting people globally, conflict-affected, displaced, and vulnerable migrant populations who were already in a precarious situation, have been severely affected by the pandemic. To ensure efforts and progress on inclusion are maintained and barriers to education for crisis-affected learners do not persist or worsen during the pandemic, disaggregated and quality data is required to ensure no one is left behind. Available of quality data also ensures that education sector responses to COVID-19, notably distance learning and the reopening learning institutions, are equitable (UNESCO, 2020). However, disaggregation by data on affected suffers through comparable challenges of inconsistency in definition of affected individuals, overlapping definitions, and the duration of crisis. For instance, for Internally Displaced Persons (IDPs) there is no consensus as to what constitutes a protracted crisis, and there is resistance to the use of duration and scale as markers of protraction (Kälin & Chapuisat, 2017).
9. Disaggregation by Language. The first or home language is defined as the student's main language of communication outside the school environment. It is usually either the first language students learn or the language of their family or local community (UNESCO-UIS, 2020a). For many students, the first or home language may not be the language of instruction. This may impact the learning outcomes of students who are not proficient in the language of instruction. Therefore, disaggregated data by language can provide clarity in targeting policies and programs to improve the learning outcomes of students whose home language differ from the language of instruction
However, when a student is asked about their respective home language, their home can have multiple languages where each parent is versed in a different language. This exposes the children to two different home languages. The EMIS metadata survey shows that most countries gather data for only one home language. Therefore, data for children who may have multiple home languages may not be recorded appropriately.

7. Way Forward

The process of data collection, consolidation, cleaning, disaggregating, and analyzing all face unique challenges across most countries. These challenges are more pronounced in low-and middle-income countries who lack adequate resources and do not have well-functioning information management systems. However, the importance of quality disaggregated data to improve education systems' ability to meet the ambitious goals of SDG 4 related to access, quality, and equity in education is irrefutable. Adopting the following recommendations can help mitigate some of the challenges that arise in education data collection and management.

Figure 12. Recommendations on a way forward for effective use of EMIS



1. Standardizing definitions. As elaborated in the previous section, most of the challenges pertaining to data disaggregation lies with the inconsistencies in definitions of terminologies used across countries and regions. Countries may have different definitions for terms such as conflict affected, internally displaced, disaster affected, internal migration, external migration, refugees, etc. It is important to develop standard definitions for these terminologies and identify how these terminologies are distinct and how they may overlap with each other. Standardization of definitions allows for comparison of education systems and quality across countries, regions, and time and limits the risk of underestimation and overestimation of data.
2. Use of Unique and Permanent ID. Lack of data interoperability, which is the seamless exchange and re-use of data across platforms and systems remains a key challenge across the world. Data is often fragmented across different systems that do not interact with each other and records are not easily reconcilable. Promoting countries to use unique and permanent ID that is lifelong and can be used across different platforms and departments can solve a lot of these challenges. For instance, if an EMIS system generates a student's ID, their education data remains within the school system. When the student leaves the school system, their student ID will not hold any record or value. Therefore, if countries coordinate with all departments and ministers to mandate the use of the same unique ID for all individuals (for example, a national ID), it can enhance interchangeability and use of data between different systems. The use of national IDs makes EMIS data more compatible with other databases which further improves the ability of ministries to make well informed policy and program decisions. Maldives is an example of a country that already uses a national ID as the student ID. However, if a child does not have a national ID due to some legal reasons, an auto-generated number by the Maldives Education Management Information System (MEMIS) is used, and for expatriate students, a passport number is used(UNESCO-UIS, 2020b).
3. Integrating the EMIS System. To address the of issue of lack of interoperability of data across systems, countries should be encouraged to integrate the EMIS system with other systems such as the health system and the social protection system so that the data across different systems can 'speak to' each other. Lack of effective coordination and communication among different departments/ministries can lead to fragmented and duplicated work and is not sustainable in the long run. Use of permanent unique IDs as discussed in the aforementioned section is one way of integrating the EMIS with other systems that can help countries with targeted cross-cutting interventions. Similarly, the metadata survey data showed that the EMIS of many countries face an issue of compatibility between software, data migration, integration, and

sustainability of the EMIS system. EMIS systems are often developed using different platforms/ architects and often software licenses are not renewed, source codes are not documented, data dictionaries are not developed or documented properly. These discrepancies hamper the integration of data from multiple EMIS into a central EMIS. Due to these reasons some countries develop an entirely new EMIS system instead of enhancing an existing EMIS system, without archiving historical data and making the new system incompatible with old systems(UNESCO-UIS, 2020b). These issues negatively impact the production of high-quality data at the national, regional, and international level and hinder the use of education data for policy and monitoring purposes.

4. Technology use. EMIS's success largely depends on timely and reliable production of data as outdated data may result in missed opportunities for program and policy interventions that are required at a given time. The traditional EMIS lacks real time date on schools, students, and teachers, to respond to urgent educational needs. The Covid-19 pandemic has caused schools to shut down worldwide and given rise to new and innovative ways of distance teaching and learning. However, the expansion of distance and hybrid learning showed that more traditional EMIS systems were not ready to reach beyond the walls of the school building to support teaching, learning, and monitoring during school closures. Therefore, the need for real-time data, individual data, support for hybrid learning systems, process-capturing (e.g., attendance), integration, and interoperability by leveraging technologies, and the inclusion of marginalized learners in data collection has been realized now more than ever (UNESCO 2021).
5. Capacity Building. Most of the nonfunctional EMIS are caused by unavailability of trained EMIS officials in the MOE and lack of an operational budget. In many cases, the nonfunctioning EMIS is led by computer professionals and lacks education planners, statisticians, and analysts to achieve the fullest utilization of EMIS functions(UNESCO-UIS, 2020b). When it comes to providing capacity building support, only the capacity of government officials is considered, whereas the capacity of headteachers and school officials responsible for producing and reporting the data is often overlooked. It is imperative to enhance capacity of all the human resources involved in data collection, production and use at national, sub-national and school levels to ensure data quality is maintained. Similarly, a human resource succession plan and proper knowledge transfer procedures must be in place to ensure that the EMIS is functional even after older staff members retire or leave. Therefore, countries and development partners need to invest in adequate financial and human resources with proper knowledge and training to make effective use of EMIS systems in planning and decision making.
6. Use of Data. It is widely recognized that use of data in education planning and implementation improves the quality of education systems. However, to increase the reliability and quality of data, data needs to be used repeatedly. Using data multiple times allows for identification of any problems with the data and provides opportunity for multiple revisions. If same data is used by multiple authorities on multiple occasions, it also helps to cross-verify the authenticity of data. If data is not used repetitively and rigorously, the quality of data remains poor. Using data creates a virtuous cycle of updating and improving data quality at every use.
7. Improve Culture of Evidence Use in Planning. In many countries, the decision-making environment is not always conducive to using evidence in planning. There often exists a disconnect between those who manage the education information system and the people involved in education planning and budgeting. Since evidence is essential to shape better education outcomes, focus should be given to provide education planners with better access to reliable and timely data. Establishing a culture of using evidence to inform planning and decision-making for education should be a priority for all governments across all levels and regions.
8. Undertake Regular Audits. For both quality assurance purposes and legal requirements, authorities need to undertake regular data

audits. Undertaking regular audits allows for identification of anomalies and in locating sources of errors. Regular audits allow to identify errors and inconsistencies in data in a timely manner, which will prevent incorrect and unreliable data from being used in planning and decision making later. UIS has produced an Education Data Quality Assessment Framework (DQAF) and the member countries can develop their own based on their context and requirements.

8. Resources at UIS

Apart from the sources listed under references, multiple resources available at the UIS were referred to compile this report. National EMIS questionnaires, websites of different countries, software adopted by respective countries and reports referred to (Operational Guide, Buyer's and User's guide, Metadata Survey and Report, Data Quality Assessment and Frameworks) were obtained from the EMIS microsite (http://emis.uis.unesco.org/). Similarly, the UNESCO-UIS Technical Cooperation Group's website (http://tcg.uis.unesco.org/) was used to access the official list of SDG4 Indicators, SDG4 National Benchmarks and dashboards for different regions and countries, Metadata and Methodological Documents. The UNESCO-UIS Global Alliance to Monitor Learning website (https://gaml.uis.unesco.org/) was used to understand the monitoring framework of different SDG4 Indicators and mapping of learning assessments to SDG4 Indicators. The UIS Statistics website (http://data.uis.unesco.org/) was used to retrieve official data on education indicators of different countries and through different time periods including government expenditure data. The Global education observatory website (https://geo.uis.unesco.org/) was referred to obtain country and region wide education related data and information to improve the monitoring of SDG4 progress, relevant and emerging policy commitments including impact of COVID-19 across countries. Similarly, forthcoming yet unpublished papers such as the UNESCO-UIS and UNHCR (2021) report on *Refugee Education Statistics: Status, Challenges and Limitations* and Scoping Education in Emergencies (EiE): A Preliminary Conceptual Framework for Understanding EiE Data were referred to obtain additional relevant information required for the report.

9. References

Kälin, W., & Chapuisat, H. E. (2017). *Breaking the Impasse: Reducing Protracted Internal Displacement as a Collective Outcome.* OCHA Policy and Studies Series. United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

UNESCO. (2011). *The hidden crisis: Armed conflict and education | Global Education Monitoring Report*. UNESCO.

UNESCO. (2015). *Education 2030 Incheon Declaration and Framework for Action*.

UNESCO. (2016). *Education for People and Planet: Creating Sustainable Futures for All, Global Education Monitoring Report*. UNESCO.

UNESCO(2020). *Why we need better data and evidence for education in emergencies*

UNESCO. (2021). *The road to future EMIS discussed at UNESCO's three-day seminar.*

UNESCO-UIS. (2012). *International standard classification of education: ISCED 2011*. UNESCO Institute for Statistics.

UNESCO-UIS. (2020a). *Operational Guide to Using EMIS to Monitor SDG 4*. UNESCO Institute for Statistics.

UNESCO-UIS. (2020b). *Data Innovation for Producing SDG 4 Indicators: A Global Analytical Report* (Information Paper No. 65; p. 51). UNESCO Institute for Statistics.

UNESCO-UIS, & GPE. (2020). Efficiency and Effectiveness in Choosing and Using EMIS.

UNICEF. (2019). *Birth Registration for Every Child by 2030: Are we on track?* UNICEF DATA (p. 44).

United Nations Statistics Division. (n.d.). *SDG Indicators-MetaData Repository*. Retrieved December 7, 2021.

USAID. (2020). *Mapping of Tools for Disaggregation by Disability Status* (p. 43).

UNSD (2017). *Principles and Recommendations for Population and Housing Censuses.* Statistical Papers. Series M, No. 67, Rev.3.

UNSD (1998). *Recommendation on Statistics of International Migration*. Statistical Papers. Series M. No. 58, Rev.1.

UN General Assembly. (2015). *A/RES/70/1 Transforming our world: The 2030 Agenda for Sustainable Development*. 35.