Education Management Information Systems

Efficiency and Effectiveness in Choosing and Using an EMIS

Guidelines for Data Management and Functionality in Education Management Information Systems (EMIS)

Chris van Wyk and Luis Crouch

United Nations Educational, Scientific and Cultural Organization

UNESCO INSTITUTE FOR STATISTICS

SUSTAINABLE DEVELOPMENT GOALS

GPE Transforming Education

Education Management Information Systems

Efficiency and Effectiveness in Choosing and Using an EMIS

Guidelines for Data Management and Functionality in Education Management Information Systems (EMIS)

Chris van Wyk and Luis Crouch

United Nations Educational, Scientific and Cultural Organization

UNESCO INSTITUTE FOR STATISTICS

SUSTAINABLE DEVELOPMENT GOALS

GPE Transforming Education

UNESCO

The constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) was adopted by 20 countries at the London Conference in November 1945 and entered into effect on 4 November 1946. The Organization currently has 195 Member States and 11 Associate Members.

The main objective of UNESCO is to contribute to peace and security in the world by promoting collaboration among nations through education, science, culture and communication in order to foster universal respect for justice, the rule of law, and the human rights and fundamental freedoms that are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations.

To fulfil its mandate, UNESCO performs five principal functions: 1) prospective studies on education, science, culture and communication for tomorrow's world; 2) the advancement, transfer and sharing of knowledge through research, training and teaching activities; 3) standard-setting actions for the preparation and adoption of internal instruments and statutory recommendations; 4) expertise through technical cooperation to Member States for their development policies and projects; and 5) the exchange of specialized information.

UNESCO Institute for Statistics

The UNESCO Institute for Statistics (UIS) is the statistical office of UNESCO and is the UN depository for global statistics in the fields of education, science, technology and innovation, culture and communication.

The UIS is the official source of internationally comparable data used to monitor progress towards the Sustainable Development Goal on education and key targets related to science, culture, communication and gender equality.

Published in 2020 by:

UNESCO Institute for Statistics
P.O. Box 6128, Succursale Centre-Ville
Montreal, Quebec H3C 3J7 Canada

Tel: +1 514-343-6880
Email: uis.publications@unesco.org
http://www.uis.unesco.org

ISBN 978-92-9189-247-1
Ref: UIS/2020/ED/TD/2

© UNESCO-UIS 2020

This publication is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (http://creativecommons.org/licenses/by-sa/3.0/igo/). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (http://www.unesco.org/open-access/terms-use-ccbysa-en).

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities or concerning the delimitation of its frontiers or boundaries.

The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

Table of contents

ACRONYMS AND ABBREVIATIONS 5

LIST OF FIGURES, BOXES AND TABLES 6

1. INTRODUCTION 7

2. ABOUT THE "BUYER'S" AND "USER'S" GUIDES 9

3. EMIS DEFINITION 11

4. EMIS ARCHITECTURE 13

5. BUYER'S GUIDE 15

5.1 The Scope and Context of EMIS 15

5.1.1 Key components of an EMIS 15

5.1.2 How the SDG 4 Indicators have changed the scope of EMIS 17

5.1.3 The Importance of a data or EMIS policy 19

5.2 EMIS Procurement 19

5.3 A Total Data Quality Framework 21

5.4 EMIS Functionality and Standards – Buyer's Guide Aspects 23

5.5 Additional Minimum Requirements and Standards Related to Technology 30

5.5.1 Software requirements 31

5.5.2 Database requirements 32

5.5.3 Telecommunications and data networking requirements (including networks and the internet) . . 33

5.6 Future Technology Considerations 34

6. USER'S GUIDE 37

6.1 Staff Requirements 37

6.2 EMIS Production Life Cycle 37

6.2.1 Compilation of the survey questionnaire (design and development) 38

6.2.2 The Allocation of unique identifiers 41

6.2.3 The Maintenance of the directory (register) of schools 43

6.2.4 Data entering 43

6.2.5 Data storage 44

6.2.6 Data interoperability 45

6.2.7 Data analysis and reporting 45
6.2.8 Data release, data dissemination and data usage 47

SUGGESTED READING 49

ANNEX A. LIST OF CONTRIBUTORS 51

ANNEX B. LIST OF SDG 4 INDICATORS AND CONSIDERATIONS FOR EMIS 53

ANNEX C. SAMPLE EMIS CALENDAR 65

Acronyms and abbreviations

ADEA	Association for the Development of Education in Africa
API	Application programming interface
AU	African Union
DBMS	Database management system
DHIS	District Health Information Software
EMIS	Education Management Information System(s)
GPE	Global Partnership for Education
HISP	Health Information Systems Program
MoE	Ministry of education
OLAP	Online analytical processing
OTP	Operational transactional processing
RDBMS	Relational database management system
SaaS	Software as a service
SDG	Sustainable Development Goal
TPS	Transaction processing system
UIS	UNESCO Institute for Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization
URS	User requirement specifications

List of figures, boxes and tables

LIST OF FIGURES

Figure 1. Components of an Enterprise Architecture 13
Figure 2. Typical Data Architecture of an EMIS 14
Figure 3. EMIS Production Life Cycle 38

LIST OF BOXES

Box 1. A Word on “Standards” 10
Box 2. Urgent Data Under COVID-19 18
Box 3. Summary of UN Principles of SDG Indicator Reporting 22
Box 4. Technical Skills for EMIS Staff at National, Subnational and School Levels 37
Box 5. School Census Approach: Manual or Web? 39
Box 6. Discrepancies Between Manual Data Collection and EMIS Input Screens 39
Box 7. Interoperability Strategy 45

LIST OF TABLES

Table 1. Dimensions of Quality of Data 22
Table 2. EMIS Functionality Standards, by Data Collection Layer 23
Table 3. Technology Upgrade Paths 35
Table B.1 SDG 4 Indicators – and How EMIS Could Approach Them 54

1. Introduction

The United Nations Education, Scientific and Cultural Organization (UNESCO) and the Global Partnership for Education (GPE) joined forces and held the first International Conference on EMIS at UNESCO Headquarters April 11–13, 2018.[1] A sentiment frequently expressed by representatives of national Education Management Information System (EMIS) units was a need for guidance from the international community on (i) how to make decisions about improvements to their EMIS; (ii) what technical basis to negotiate with vendors or suppliers; and (iii) how to better discuss their requirements with development partners wishing to contribute to the development of their data for education, both lacking in comparative information about what a modern EMIS ought to be able to do.[2] Unlike in the health sector, where a default standard system has emerged through the District Health Information Software (DHIS), in education there is a proliferation of systems proffered by various donors and suppliers. In addition, ministries of education lack clarity as to how other data systems (e.g. assessments, teacher HR) should interface with EMIS. In line with this, a need was also expressed for more unified ways to diagnose EMIS as well as a more standardized default EMIS platform (perhaps with a view toward education in emergencies). This guide – essentially a "Buyer's Guide" and a "User's Guide" to EMIS – is an attempt by UNESCO's Institute for Statistics (UIS) and GPE to address these needs expressed by countries, and is the result of numerous discussions in various forums leading up to and including a meeting of the Data Solutions Roundtable (DRT) in September 2019 at the margins of the UN General Assembly.[3]

Over the past decade, a plethora of systems for collection, management and reporting of education data has evolved. Consequently, deciding what system to use is a difficult and challenging task, as a system's capabilities have to align with a country's objectives and priorities. The aim of this report is to assist countries to make a more informed choice in obtaining specific and relevant EMIS software solutions. As the name implies, this report is a guide (or manual) – it's objective is not to lay out philosophies (unless absolutely necessary), report on research findings or engage in advocacy. This guide is intended for a technical audience (i.e. data, IT, planning) in developing countries or in the international agencies and academic institutions that work with them. Nonetheless, nontechnical policymaker audiences in a ministry may benefit from some of the early sections in particular.

Effective decision-making relies on quality data managed in efficient information systems. Information is a necessary resource, produced by information systems, and is a key building block to management and decision-making in education. An EMIS is not only a technological solution restricted to operational processes. As an information system,

1 See https://en.unesco.org/themes/education-policy-planning/emis-2018.

2 See https://en.unesco.org/news/unesco-gpe-launch-first-international-conference-education-management-information-systems-0. Also noted was a need to harmonize the various frameworks used by international agencies (e.g. World Bank, UNESCO, UNICEF, ADEA) to evaluate EMIS. While this is a valid and important need, this guide focuses on the *systems* per se.

3 https://www.globalpartnership.org/content/outcomes-education-data-solutions-roundtable – see Outcome 4.

it should facilitate strategic decision-making, policy formulation and budgeting, and, if possible and relevant, routine management at at district level, in helping schools. In general, however, this report does not address school-level data recording systems for internal school management.

An EMIS is typically a function and unit within a country's ministry of education (MoE), with the responsibility to acquire and maintain an integrated education information system for the management of education. A more complete definition of an EMIS is provided in Section 3.

EMIS has a responsibility to collect, process, analyze and disseminate education data and information in order to support the monitoring and evaluation of the performance of the education system. Whether the EMIS Unit carries out the analysis stage or whether this is done in offices such as planning, budgeting or policy, will differ from country to country, and should ideally be based on where the right skills profile exists (refer to Section 6.1 for details on required technical skills of EMIS staff). At the very least, the EMIS unit can provide some value-add as a service to other units, even if it is only to produce relevant indicators for them.

2. About the "Buyer's" and "User's" Guides

This guide is a collective effort and has been developed by an EMIS consultant appointed by the UIS.[4] The development process was informed by a literature review and inputs from key role players from existing and evolving EMIS platforms. The guide comprises two main sections:

1. A "Buyer's Guide" that refers to the standards of functionality a system ought to have (Section 5).

2. A "User's Guide" that refers to how to make better use of an EMIS once a country has it (Section 6).

Although a MoE would never strictly "buy" an EMIS – indeed, an EMIS is more akin to an ecosystem within which one operates and links many units of an education ministry, other ministries, academia and civil society – many of the decisions to be made when procuring or improving an EMIS resemble a purchasing process. The "Buyer's Guide" – a description used here loosely – sets out the standards of functionality for an EMIS, regardless of the actual method of acquisition or development. (See Box 1 for what we mean by "Standards" as something that is not prescriptive.)

The assumption here is that most ministries would procure (or be provided by a development partner) new additions to an existing "EMIS backbone",[5] or that they would use off-the-shelf (mass produced) tools to replace or add new features. Occasionally a ministry might turn its EMIS over completely. Furthermore, because at least *some* aspects of improving an existing platform (or very occasionally the development of an entirely new platform) may be procured by a ministry or procured and provided by a development partner, this section includes guidance on procurement. After "buying" a system, one procures it, and then engages in a process of using it – hence the "User's Guide" that follows. Again, this description is used loosely, and more as a narrative device.

As background work, this report researched various EMIS and related software platforms to compile a list of functionality standards and used reasoning based on typical reporting requirements, including the Sustainable Development Goal (SDG) 4 indicators, as well as regional variants such as those selected to monitor the African Union's Continental Education Strategy for Africa (CESA 16-25) – but without giving primacy to the latter, as country needs should be paramount. The aim was to develop a set of standards to illustrate what the EMIS must be able to do, in order to supply accurate, valid information to education sector policymakers and school managers and international organizations to whom most countries

4 Chris van Wyk (Stellenbosch University, South Africa), chrisvanwyk@sun.ac.za. The development of this report was contingent on the cooperation, consultation and input of representatives of organizations of existing and evolving EMIS platforms. Representatives of these organizations have been more than helpful, and sincere thanks are extended to them for their assistance. Collaborating on the draft was Luis Crouch (RTI International, individually consulting to GPE). Contributors who drafted small sections or otherwise helped are listed in Annex A.

5 For a definition and discussion, see Section 3.

Box 1. A Word on "Standards"

This guide repeatedly uses the word "standards." This is not meant as an imposition on countries' own needs. On the contrary, these standards are meant to be helpful to EMIS units in offering a sense of what other units in other countries may be doing (or aspire to be doing), and what experts think may be a good idea. There is no suggestion, however, that any system has to meet all these standards in order to be considered a good system. There are many local adaptations that are possible and indeed necessary. Two examples include:

Level of decentralization in each country. In a truly federal state, the role of a central level in setting norms may be quite different from those in a unitary nation, even one that is extremely decentralized. For instance, a federal state may have more limited powers in asking its member states or provinces to implement a common software platform. And in any case it may not be optimal as needs may vary. Options for federal states are discussed throughout the document as a way of illustrating these differences and how these standards can vary. In a unitary state that is decentralized, the governance structure for EMIS may consider a committee of provincial heads of EMIS to make joint decisions. An approach like this may be less necessary in a centralized state.

Baseline education situation in each country. Countries may prioritize the intensity of effort for different indicators. For example, in a country where *access* is still a major challenge, access indicators may need to be more granular or refined and may need to have more links to databases that can help determine why access may be low.

owe reporting. The description of how various existing platforms relate to the standards laid out in the User's Guide below (Section 6) is tentatively contained in a *separate* document (not an annex) titled "Education Management Information Systems: Software Platforms."

3. EMIS Definition

This guide relies on the following definition provided by UNESCO:[6]

> *"An EMIS can be defined as 'a system for the collection, integration, processing, maintenance and dissemination of data and information to support decision-making, policy analysis and formulation, planning, monitoring and management at all levels of an education system. It is a system of people, technology, models, methods, processes, procedures, rules and regulations that function together to provide education leaders, decisionmakers and managers at all levels with a comprehensive, integrated set of relevant, reliable, unambiguous and timely data and information to support them in completion of their responsibilities'."* (UNESCO 2019)

This definition states, or strongly implies or assumes, that

- EMIS is more than just a technical solution;
- Education has an integrated and distributed nature;
- All levels (tiers) of the education structure should be considered (institutional, subnational and national); and
- All systems and technical infrastructure should be in place.

In what follows, "EMIS backbone" will refer to the set of processes and data that comprise the classical administrative counts (such as enrollment, teachers, supplies) created via annual (or periodic) censuses and tallies of school data, sometimes augmented by population data or other similar data needed to create the access and flow indicators of an education system (such as gross enrollment ratios, net enrollment ratios, estimates of completion ratios for each level). "Data warehouse" (used interchangeably with "data mart") will refer to the set of data that can be added and linked to the classical EMIS backbone data to produce important value-add analyses, whether these are done by the EMIS office or by other user offices such as planning, budgeting, policy analysis, teacher training, health or poverty.[7] To function properly, these databases should be linkable via school IDs, district IDs, etc. Whether for the EMIS backbone itself or an expanded warehouse, automatic data collection, such as from a learning management system or administrative data collected at school level, should be considered, though for the lowest-income countries this is more likely a future orientation (for more futuristic scenarios, see Section 5.6.). Finally, there are many important indicators that are often binary or qualitative that do not properly belong in the EMIS backbone or even on the data warehouse, such as whether a country possesses a pro-

6 Though experts may disagree on how to define an EMIS, this definition is consistent with the aims of this guide.

7 There is no hard-and-fast distinction between what constitutes the traditional EMIS backbone and a broader, related data warehouse that is useful to the ministry. Classical, school-by-school enrollment counts belong on the backbone. Learner performance data based on random sample assessments or public exams typically belong in a broader warehouse. But there are variables such as population in the catchment area of schools that might be used so repeatedly that they could be considered part of the backbone.

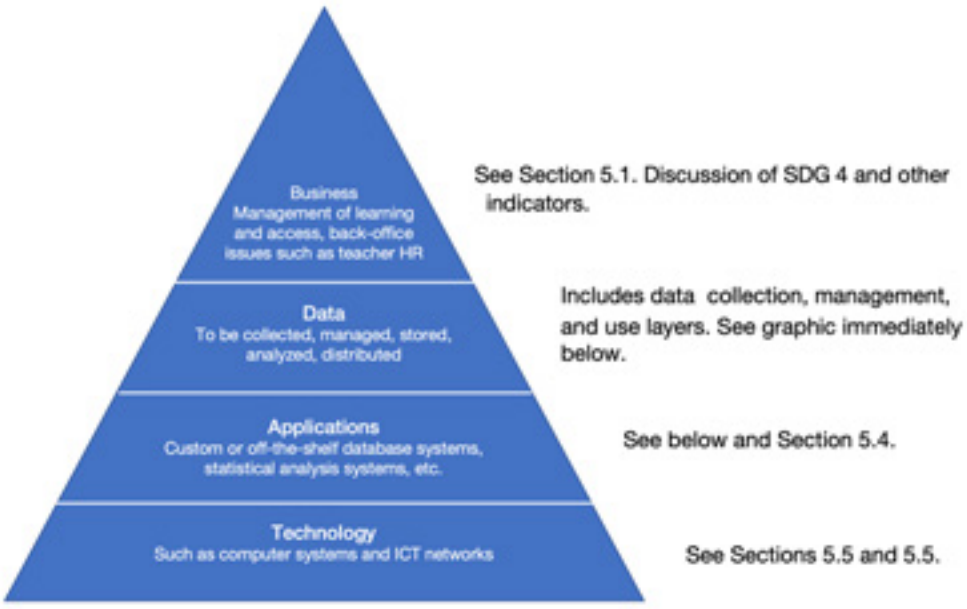
poor school funding system or whether there is a curricular policy on X or Y. While many stakeholders will want to know such things, and while these are often reporting obligations of the ministry (including on SDG 4 – see Section 5.1.2), they are optimally maintained or gathered by not an EMIS but a planning or policy office of the ministry.

4. EMIS Architecture

In simple terms, the EMIS architecture underpins the way in which data are collected, stored and reported, and lies at the heart of the Buyer's and User's Guides. The components of any software solution include *data collection*, *database management* and *utilization* (which includes *analysis* and *reporting tools*). The Buyer's and User's Guides draw on the EMIS architecture discussion in this section, wich provides the main content and structure for these sections. In general, the EMIS architecture (or of information systems or for any organization) should respond to the overall *enterprise architecture* – in this case, that of a typical ministry. Figure 1 represents four components of an overall enterprise architecture to be considered, namely:

1. *Business architecture*: the domestic and international needs/requirements of the MoE – briefly covered in Section 5.1 (see the discussion on SDG 4).
2. *Data architecture* – covered below.
3. *Application architecture* – covered below but also in Section 5.4.
4. *Technology architecture*: based on the choices/needs in levels above, technologies are chosen – covered partly in Section 5.4 and elaborated on briefly in Section 5.5.

Figure 1. Components of an Enterprise Architecture



Source: Authors.

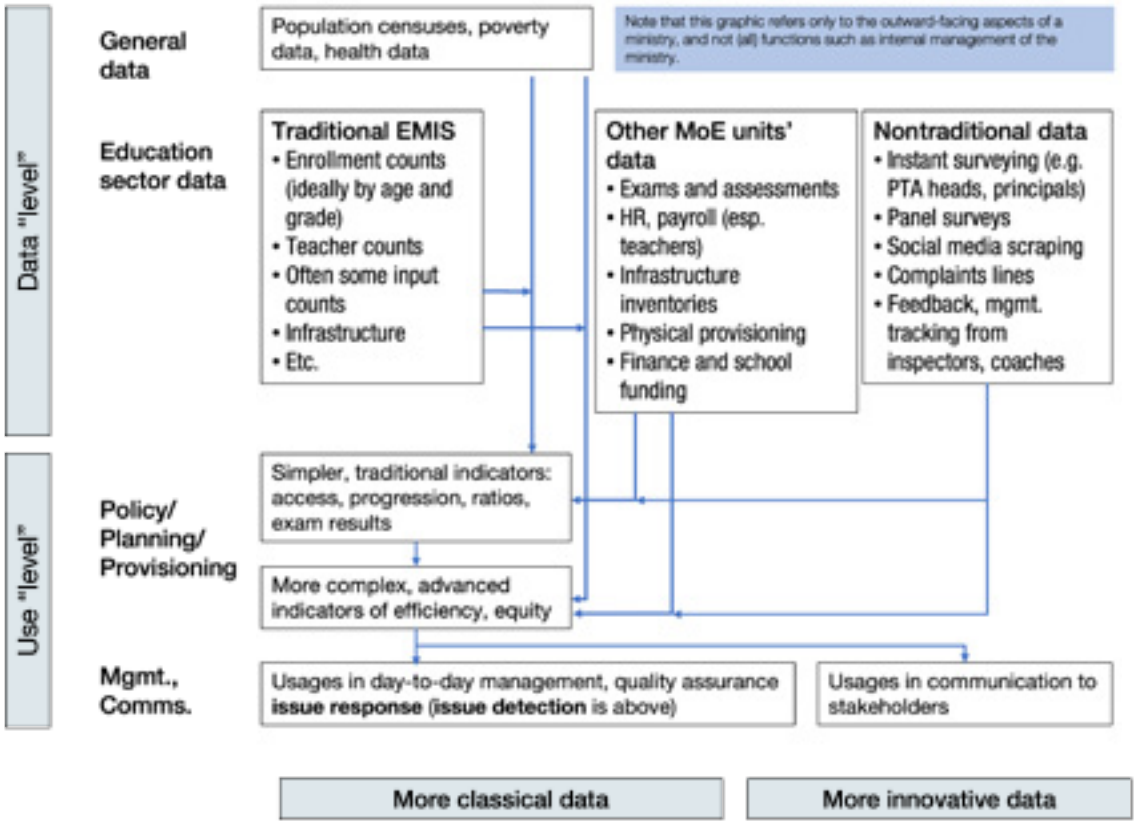
The following is a summary of the most common components of any *EMIS data architecture and guides:*

- *Data collection layer*: how data are gathered.
- *Data management and storage layer*: how data are managed and stored in the data mart/ data warehouse as well as transferred between layers in the system if there cannot be true distributed entry and management.
- *Data utilization layer*: how data are analyzed, reported and disseminated.

Figure 2 shows how these are related in the data architecture, and what the sources of data typically are.

The following sections provide a more detailed overview of the most common components of any EMIS architecture and its related functions, activities and processes, from data collection to data processing and storage to data analysis, reporting, dissemination and use.

Figure 2. Typical Data Architecture of an EMIS



Source: Authors.

5. Buyer's Guide

5.1 The Scope and Context of EMIS

Education systems are, by nature, extremely complex and multifaceted, and the need for data as well as for evidence-based planning, monitoring and evaluation, decision-making, budgeting, policy formulation, resource allocation and management can be similarly complex and multifaceted. Furthermore, all education systems are different, though they share many features (for example, it is hard to think of an education system – at least for now – without teachers, students, schools, administrators or textbooks, or that does not need funding). Thus, one can conceptualize of a relatively common data architecture, but always with the need to adapt to the particular ministry's aims and demands.

EMIS is not optimally designed and mandated to produce *all* the datasets needed to meet the requirements of the education sector. It ought to, however, provide the analyst with "hooks" with which to integrate other data (e.g. poverty, population), even if only ex post and in a process of value-add that is not part of the actual EMIS backbone. Data interoperability (ability of various databases to link to each other, for example, via common school IDs) is an important principle in the production of quality data and a key element in overall architecture of EMIS – it is really the main source of value added by an EMIS to the higher levels of the ministry. Furthermore, EMIS should be a system that can respond to the present demands but also evolve according to emerging and future demands and technical possibilities. In some cases this could be a service provided by the EMIS unit in creating a data warehouse. Furthermore, an EMIS should not be thought of as a technological solution alone and restricted to the operational processes only. Therefore, EMIS should not try to encompass all possible forms of data collection processes, as this can easily become unmanageable and risks overloading staff. EMIS simply cannot be the one-stop shop for all the data collection processes required in the education sector; i.e. it cannot be everything to everyone. A strong EMIS should, however, be able to link to everything.

5.1.1 Key components of an EMIS

This section describes the context and scope of an EMIS and also describes the education administrative data collection processes, procedures and challenges during the development of an integrated EMIS. Administrative data here refers to data collected from schools by the national office for recordkeeping, planning, monitoring, evaluation and policy formulation. The basis of EMIS should be data for schools, learners, teachers and physical facilities. A complete EMIS should ideally include data on enrollments, attendance, completion rates, learning assessments, student health, finance, teacher characteristics, and some administrative statistics.

For the purpose of this guide and according to aggregated data collection systems (i.e. datasets typically limited to centralized databases at school level), EMIS should include among others:

Basic enrollment and input counts:

- Enrollment (by grade, gender, poverty, age, subject, disability, linguistic groups, etc.)

- Repetition (by grade, gender), dropouts, completion, flows in general
- Learner demographics
- Teacher (experience, qualification, gender, placement, appointment, etc.)
- School (physical facilities, general information, location, classrooms, etc.)

Other components within the education sector:

- School feeding program (where applicable)
- Library and archives services
- School hostels (where applicable)
- Learner health
- Learning materials (textbooks at a minimum, typically)
- Performance data (examinations and learner assessments – ideally as an intrinsic part of the EMIS; at a minimum the EMIS needs to be able to link seamlessly to assessments)
- Teacher qualification and salary (links to both HR and payroll)
- School funding (in many countries schools receive direct funds for certain expenses) and budget

Other line ministries:

- Higher education (if a different ministry rather than a level within the MoE)
- Labor/employment (typically training programs at adult or higher education levels)
- Home affairs, or foreign affairs (international student flows, visa issues, etc.)
- Health (student health statistics)
- Social services (student family grants and poverty data if relevant)
- Finance or planning (poverty data for targeting, for example)

Other organizations (most likely not as part of the EMIS itself, but enabling links if possible)

- Non-government organizations (NGOs)
- Population data (e.g. national statistics office)
- Municipalities (where they are relevant to education)

All of these ought to be available by level (from pre-primary and early childhood development/ECD on up) and by administrative jurisdiction (school, district, province and national, as appropriate for the country in question).

Other datasets collected by other components within the education sector and other line ministries are equally important for evidence-based decision-making. While it is impossible to provide a comprehensive list as this will depend on a country's priorities, they typically might include items such as the variables related to SDG 4 that a country EMIS may want to produce; indeed, using SDG 4 as a case in point can provide a useful default list. Another useful reference is a list of harmonized indicator goals set under Africa's continental education strategy, CESA, with SDG 4 (AU and ADEA 2018). Some of the interesting issues that these frameworks raise are covered in the following section. In addition, in Annex B we list the indicators that most countries are, or should be, committed to providing under SDG 4, as reporting on these is an international commitment acquired by each country's government; this compilation presents a well thought-out and logical list of indicators, manyof which should be suitable for many countries. It is imperative to point out that this guide should not be misinterpreted as a guidance tool on how to serve donors and international agency requirements. See Box 3 on SDG reporting that reiterates that country needs are paramount.

5.1.2 How the SDG 4 Indicators have changed the scope of EMIS

The main challenge represented by the SDG 4 indicators is that many (in reality just about all) require the use of databases that go beyond the EMIS backbone of data. This was true also to a large extent during the preceding period, in which the Millennium Development Goals (MDGs) were a focus. However, the contrast between what is hoped for and what currently exists is rather more striking today. The importance of creating warehouses that have data collections that go beyond the traditional EMIS data backbone is becoming increasingly clear. For EMIS to respond to the increasing demands, it must operate as an integrated and interconnected system with datasets coming from diverse sources. A responsive EMIS would work on data inputs, processes and outputs at the different levels and modalities of educational provision with a high degree of granularity; i.e. from aggregated data at national level to the detail of a particular school, classroom or even individual student (while recognizing privacy concerns in line with national and international cross-sectoral and education standards[8]). In addition, the SDGs require disaggregation according to factors that may not always be present in an EMIS backbone, such as wealth. In some cases, it may be necessary to link the EMIS backbone data to, say, a detailed poverty map of the country. It is also fairly clear that while even simple reporting presents certain challenges, the real challenges in data are in *using* data to drive improvements on the SDGs.

Annex B lists the indicators that most countries are, or should be, committed to providing under SDG 4, as reporting on these is an international commitment required by each country's government, and how EMIS could approach these. It is imperative to point out, however, that this guide should not be interpreted as a tool on how to serve donors and international agency requirements. On the contrary: *country needs* are paramount.

Below are three examples of how SDG 4 might affect data collection and reporting:

Learning outcomes: the need and opportunity to track learning outcomes (SDGs 4.1.1, 4.2.1, 4.2.3, 4.4.2, etc.). EMIS units traditionally do not collect data on learning. But data on learning could be warehoused near the EMIS in order to see, for instance, how strong a correlation there is between learning data and measures that EMIS usually can produce, such as profiles of grade progression in schools, in order to have early warning of problems in learning outcomes. This would require making better and more automatic linkages across datasets.

Population (almost all of the access SDG sub-indicators, such as 4.1.3, 4.1.5, 4.2.2, 4.2.4). There is a need to better understand population data and how they interact with enrollment data. For example, comparisons could draw attention to flow inefficiencies and completion rates stuck at very low levels (e.g. 65%), which in turn could be a proxy for low learning, and could highlight the fact that there may be much more repetition than dropout in the foundational grades or might be helpful in triangulating weaknesses in the repetition data. This requires that EMIS staff be able to access population data that are as detailed as possible, though noting that population data become less and less reliable, when being compared with enrollment data, the smaller the

8 Privacy and security standards are a minimum for countries' ICT ministries (or equivalent). An MoE may want to go further and consider how these standards apply to data that typically only the education sector would have, such as individual children's cognitive development.

unit of analysis. Here, the ability to link to other datasets such as a detailed poverty map is key, because in most countries it will not be possible to estimate access by poverty level, for example, without linking to such a dataset.

Language and language diversity issues (SDG 4.5.2). Schools where there is a lot of linguistic diversity may require special attention, especially in the foundation years and in situations where none of the children speak the language of instruction at home. Yet EMIS traditionally do not collect these sorts of data. In this case, the issue is not so much being able to link such data to other variables (such as learning outcomes), interesting as it may be, but simply to produce the data. Options include piloting a small change to the instrument that could collect those data, as a contribution to the improvement of EMIS, or, in an approximation that is less accurate but also much less costly, linking to datasets from a census that contains linguistic information.

Box 2. Urgent Data Under COVID-19

In June 2020, schools were emerging from nationwide closures due to the pandemic created by the novel coronavirus SARS-CoV-2, with some remaining closed. In light of this, the UIS published guidance on essential data in these circumstances,[9] which emphasized the need to maintain essential or minimal data on

- Student participation in all e-platforms (or distance learning more generally) of education delivery disaggregated by individual student characteristics such as gender and poverty;
- Teacher participation in all e-platforms (or distance learning more generally) of education delivery disaggregated by individual teacher characteristics such as gender and contract status; and
- Use of short and quick tests for the frequent measurement of student learning.

Learnings from this situation could apply to broader issues around other emergencies, refugees and internally displaced persons.

9 See, for example, the fact sheet "The Need to Collect Essential Education Data During the COVID-19 Crisis" at http://uis.unesco.org/en/news/fact-sheet-essential-education-data-collection-during-covid-19-crisis.

Datasets such as those listed above will remain fragmented islands of data and will exist in isolation if they are not linked. As mentioned earlier, data interoperability is an important principle in the production of quality data and a key element in the overall architecture of EMIS. In fact, data interoperability is the only way to allow end users including policymakers to understand the significant value-add EMIS data offer in tackling key challenges. An awareness of this potential value-add can in turn increase the demand and support for the EMIS unit from planning, school funding and other user units that feed directly into the ministry's top levels. An EMIS needs to offer more than just "management" of data as its primary function. Nationally centralized data are often hard to use for true day-to-day management, and an EMIS that offers the capacity for data interoperability and provides "hooks" to the units that can create value-add is more likely to

generate greater demand for its services and hence the budgetary support of the ministry.

5.1.3 The Importance of a data or EMIS policy

A legal or policy framework should exist mandating EMIS as the custodian of education data. This would allow EMIS adequate data sharing and coordination with other data collection divisions within the education sector as well as other data collection processes outside of EMIS and the education sector. Data integration of all these datasets must be one of the most important EMIS development strategies. In terms of the "Buyer's Guide," for instance, it should be clear that without a guiding policy framework, it is difficult to know what one ought to "buy" (or build or receive in donation). Specific items for the policy to consider include the following two, which are by no means the highest priority for the policy but are relevant for determining user features:

- EMIS at a national level should take the lead to develop and implement a data sharing and data integration policy within the country as well as beyond the country. This can include issues such as privacy and security, duties, ethics, rights and responsibilities for data.[10] Some specification of budgetary and staff needs can be included in such a policy.
- In order to facilitate data integration, data sharing and exchange of data, the national office should establish a formal working group or task team to clean, link and maintain data standards. Such a group could consist of thematic leaders from the national level and EMIS leaders from relevant subnational levels to develop protocols and agreements for the linking of and maintenance of common unique identifiers for education-related datasets. The group should also ensure that this is taking place, supervising the process, as well as hold each other accountable for respective responsibilities and commitments. The actual work would typically be performed by their technician subordinates.

5.2 EMIS Procurement

The procurement of IT technology and services refers to the acquisition of technology and technical tools that align with the needs of the EMIS unit. EMIS procurement is the process of acquiring information system infrastructure (assets) to drive the EMIS operations. Note that, as explained in Section 2, in almost all cases a ministry would not go into the market and procure a whole new system. Instead, ministries may procure certain aspects, or be donated certain aspects by development partners. Even in the latter case, being clear about what the ministry wants, why and on what terms is crucial both for efficacy and budget (as donations too have opportunity costs).

The procurement vision should line up with the rest of the Buyer's Guide and the User's Guide recommendations or standards. Below are the starting points for the specifications needed, after a more general preamble.

Procurement of information systems has to be seen with a long-term, life cycle perspective. Modern systems built as progressive apps for web and mobile devices can run on a range of hardware (the computer equipment used to perform input, processing and output activities – see more in

10 For a typical set of concerns from a high-income system, see https://professionallyspeaking.oct.ca/march_2008/privacy.asp. This site also has further references to government and security privacy policies with regard to U.K. policy: https://www.gov.uk/government/publications/esfa-privacy-notice. For a discussion in the context of an emerging economy (South Africa), see https://www.isasa.org/the-protection-of-personal-information-act-and-south-african-schools/.

Section 5.5). Apps perform better with more recent browsers and up-to-date mobile operating systems, and do not necessarily require high-end hardware, as most of the heavy computation typically takes place on the server. Investment in good connectivity is arguably more important than the latest hardware.

A national EMIS is a complex system, with potentially tens or hundreds of thousands of users in the long term, if one takes into account not just administrators and planners, but also teachers, and even communities and parents who might need access to some of the information. While technology makes rapid progress, the demands on the system in terms of complexity are bound to increase, from various stakeholders. It therefore becomes essential to also consider the support structure within the organization. This is to some extent technical (managing servers and devices), but predominantly organizational, in the sense that information is the backbone of the education sector. Reaching the full potential of the investments made requires not only the high cost of procuring thousands of devices and paying for internet connectivity, but also training and competence building within the sector in the use of data for action. Therefore, an important consideration for procuring any software solution is the long-term community of practice around the system; i.e. people who focus on improving the collection and use of education data for a range of purposes. Procuring a national system is not a matter of, for example, simply offering a contract for 2,000 laptops to ensure all schools have access (though such a purchase is likely necessary), but having a plan both for training in core functionality and for future expansion as the organization learns about what is possible and where the greatest benefits could accrue, based on a combination of its own experience and learning from others in the international EMIS community. A procurement vision therefore needs to be holistic: the overall process should be regarded as long term, while specific contracts for hardware can be handled in a similar manner as other government procurement, with public tenders. It may also make sense for the education ministry to coordinate with ministries in other sectors such as communication, health and agriculture. Similarly, support for the system can often benefit from being handled by a cross-government data center or national cloud, and involving the private sector, such as telecom companies or other service providers and vendors, as properly and securely managing servers often strains the capacities of line ministries.

Elements to consider in procurement include the following:

- **Operating in line with internal systems procurement process**
  - Securing guidelines and sometimes approval of the procurement section of the ministry, or the state-wide procurement if needed.
  - Defining the IT requirement (that is, the IT should be able to handle the requirements in this guide – also note that certain IT options may not be viable if one has to add to a legacy system).
  - Establishing discussions with the procurement experts from the ministry or elsewhere in government, from the moment that IT procurement becomes an objective. This should be based on a user needs diagnostic or specification.
  - Managing vendors – thought should be given to how much one can and should involve vendors in understanding the requirements from early on (they may be helpful but may then acquire information others do not have access to; follow government procurement guidelines).

- Verifying the quality of products and services to be provided by asking for references, past performance, both in terms of process and final satisfaction of other consumers.

- **Differentiated aspects of procurement according to what is being procured; there may be different vendors for different aspects**
  - Hardware procurement: hardware includes servers, routers, computers, printers, etc.
  - Software procurement: software includes licensed software, in-house developed software, etc.
  - Services: IT infrastructure is more than computers and applications and includes networking infrastructure connecting different computers, printers and servers internally and to the web.

- **Ongoing support plans and costs**
  - "Life of product" costs and assumptions about technical requirements that the ministry should be ready to take on. This should be done with reference to staffing needs and plans. Training of staff could be part of IT procurement, both short and long term. Note that open-source software may have low up-front costs but high life of product or other hidden costs.

- **Procurement risks**
  - Time frames for procurement – it is important to be realistic, such as not attempting to procure something very expensive in 10 days; seek general guidelines from ministry or general government procurement.
  - Funding/budget – the budget must not be unrealistic (insufficient) or it may encourage vendors to lowball or make unrealistic promises.
  - Time frame for performance should be realistic or vendors may make unrealistic promises to get the work done; consult with IT experts.
  - Not anticipating life of product costs and technical requirements on the ministry.

5.3 A Total Data Quality Framework

All functionality standards have to end up supporting some aspect of total data quality. To ensure data quality, national EMIS needs to establish a *data quality system*, complete with a data quality approach to standardize the data, provide definitions and establish data naming conventions. This is especially useful when data are transferred from one level to another, for example, from national to subnational to schools or from subnational to national.

Such a system should identify quality requirements and criteria; key EMIS processes and their sequence and interaction; the policies; and the criteria and methods for defining, detecting, correcting and preventing nonconformity.

The quality of data can be defined in terms of *completeness*, *relevance*, *accuracy*, *timeliness* and *accessibility*. These dimensions of quality are described in Table 1.

Another useful reflection on principles that determine quality (as well as other issues such as usability, sustainability, etc.) are the UN's "Principles of SDG Indicator Reporting and Dissemination Platforms and Guidelines for Their Application" (see Box 3 for a summary).[11]

11 For full explanations, see https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3a-Principles-guidelines-SDG-Monitoring-Reporting-Platforms-E.pdf.

Table 1. Dimensions of Quality of Data

Dimension	Description
Completeness	Data for all data elements for all reporting entities (e.g. schools, learners) should have been submitted. Data are complete when data values are present for all records, occurrences or logical entities in pre-primary, primary and secondary education, including learners with special educational needs.[12]
Relevance	There is no point in collecting data unless it is put to some meaningful use in education decisions. Excessively long questionnaires should be avoided.
Accuracy	Accuracy should be prioritized in the data capturing methods, data validation methods and data verification processes, possibly at the expense of time and money. Data should be within the normal range for data collected for that specific data element and entity. Check for outliers – the capacity to check for automatic detection of most outliers upon entry should be part of the user specifications.
Timeliness	Data should be current or up to date. Data must be on time and available when they are required, otherwise the credibility of the information system diminishes. Data from all the reporting institutions should be submitted at the appointed time.
Accessibility	Data should be accessible to the users at all levels of the education system. The user should know what data are available, as well as where to find and retrieve the data. Metadata and data dictionaries are important to enhance the accessibility of data within an organization. Data should be available when-needed, as-needed in a manner that is as open as is consistent with broader privacy and security concerns as well as with the country's multisectoral data policy (see Footnote 11).

Source: van Wyk 2006, modified by authors.

12 Note that these descriptions are fairly present oriented and most appropriate for typical situations now in low-income countries. A more futuristic orientation might have slightly different descriptors – see Section 5.6.

Box 3. Summary of UN Principles of SDG Indicator Reporting

- Clear institutional arrangements and management of data
- Fitness for purpose
- Dissemination platforms compliant with fundamental principles of official statistics
- Sustainability
- Interoperability and statistical standards
- National ownership
- Collaboration
- User-centered design
- Emphasis on data communication
- Data disaggregation
- Modularity and extensibility
- Standardized interfaces
- Metadata availability
- Open data
- Linked data

5.4 EMIS Functionality and Standards – Buyer's Guide Aspects

In Table 2, some functionality standards are features that a system must possess, based on the authors' and reviewers' experience with EMIS and the EMIS literature; others are more open-ended choices with pros and cons. The technology a system should possess or be based on is also a crucial consideration, and is briefly elaborated in Section 5.5.[13] The table refers to the data architecture of a typical EMIS, as referenced in Section 4.

13 For a discussion of the new issues created by the SDG 4 indicators, see Section 5.1.2.

The following are three conventions used in this section to highlight important information (i.e. they are not asides):

- **Note:** more detailed and additional information to explain the functionality that should be considered.
- **Tip:** useful advice, such as how to perform a particular task more efficiently
- **Caution:** a warning about a choice or decision that the "buyer" needs to think about

Table 2. EMIS Functionality Standards, by Data Collection Layer

DATA COLLECTION LAYER (AGGREGATED)		
Feature	**Description**	**Essential points to consider**
NOTE:	**Aggregated vs. individual-level data:** Aggregated data refers to data collected at the school level (the school census approach). Aggregated or summary data collection provides information on issues at institutional level, such as enrollment by grade, age or gender. Unit record (individual) data refers to the data collected for each learner through a school administration and management system and is discussed after this section.	
Directory	The system has a feature to **manage a directory of institutions**.	
	The system has a feature to **add and delete institutions** (i.e. to maintain the directory to stay up to date).	
	The system has the capability to **import institutional information** from other systems (e.g. Excel, csv).	
Unique Identifier	Allocate unique identifier for **institutions** according to specific algorithms.	**Tip:** The Master File System for institutions is a key process to provide a unique identifier to an institution and should be supervised at the national level. This may be easier or harder depending how whether the system is truly federal as opposed to merely decentralized. **Tip:** The unique identifier could contain the geographical codes of the governance organizations above the school so as to make it easier to link to databases (e.g. poverty) that do not go "down" to school level. Alternatively, the file should contain fields that denote the organizational unit the school belongs to (e.g. district, province). **Tip:** Ideally a good system should contain the ability to generate (possibly with a manual assist) new codes from a look-up database and a set of rules. For instance, it could assign a school an ID code based on the district and province it is located in. **Caution:** Avoid using a sequential number; rather use a specific algorithm that includes codes for subnational levels.

DATA COLLECTION LAYER (AGGREGATED)		
Feature	**Description**	**Essential points to consider**
Questionnaire Design	The system enables countries to **design custom questionnaire** forms to be used at the institution level (the school census approach).	
	The system is **customizable and flexible** to add fields according to the use case. For instance, it should allow a new field to account for, say, a new category of teacher that had not been considered before.	
Baseline Data Transfer	The system has the capability that enables users to **transfer existing data** from previous years into a **new database** for all institutions.	**Note:** The system has the capability to store data across multiple years, thus enabling the seamless transfer of the information from one year to the next. **Tip:** This ability is key to the generation of time trends and other forms of value-add data. **Tip:** Where applicable, the system should have the option to preprint (or pre-populate if online) baseline data on the physical questionnaire before it is disseminated to the institutions. **Caution:** Preprinting (as opposed to online pre-populating where possible) vastly increases the logistical requirements as exactly the right questionnaire has to get to exactly the right school.
Barcoding	The system has a **barcoding feature** to manage the dissemination and collection of questionnaires.	**Note:** This is only applicable where questionnaires are physically printed and disseminated to schools.
Data Entry	The system has the capability to set up **data entry at any level** (national, subnational and institutional).	**Caution:** *Centralization vs. decentralization* – in practice, the decentralization of systems involves the transfer of authority from the national office to the subnational offices. While acknowledging the significant advantages of decentralized information systems, it is also necessary to be aware of the risks and disadvantages. Given that the national office is further removed from the detail and often unaware of the decisions made at subnational levels, the national office should ensure that it does not decentralize its responsibility and consequently lose control of the norming and supervision of the data collection processes.
Data Entry: Technology	Data entry takes place on a **computer keyboard locally** as a manual process.	**Note:** This functionality exists on a local computer without connectivity.
	Data entry takes places over a **browser** using internet protocols. Could involve manual processes in a browser or automated processes with application programming interface (API) calls.	**Note:** This is especially advantageous in a decentralized EMIS. **Caution:** Ensure that the bandwidth and connectivity adhere to system requirements. **Caution:** Ensure that for areas of the country where bandwidth is not completely reliable and of high speed, that the system can take asynchronous data entry in any reasonable device, such as tablets. **Note:** An API is software that can call data from other databases that are programd with a different software, such as a cell phone's contacts app being able to pull up a list of school principals from the EMIS.
	The system has the feature to enable organizations to develop and deliver content to **mobile devices** and takes advantage of mobile devices' native capabilities (e.g. a contacts app and that app's ability to use either data or Wi-Fi).	
Data Entry: Quality Assurance	The system has **built-in quality assurance rules**, such as authentication and validations.	**Note:** The system allows the user to define various validation rules to identify violations. At the point of data entry, a check should be made to see if the data fall within acceptable range levels of minimum and maximum values for any particular data element. **Tip:** The system should automatically calculate certain ratios (not necessarily to put into the form) and use those as error traps, as error ranges for ratios (e.g. the pupil-teacher ratio) are easier to set up than absolute value ranges. **Caution:** Be aware that the built-in validation and verification rules do not slow down the data entry process. Furthermore, the validation rules should alert the user but must not stop the user from continuing with the data entry process.

DATA COLLECTION LAYER (UNIT LEVEL)		
Feature	**Description**	**Essential points to consider**
Note: Learner unit record data refers to the data collected for each learner through a school administration and management system. (It may apply to teachers as well of course; here learners are used as the key example). **Caution:** Should a unit record system be developed and data stored at learner unit record level at the national level? There are specific ***privacy***, ***security*** (on which, see below), ***technical*** and ***capacity*** challenges relating to this question. For example, taking into account the number of learners in a country, one database table with millions of learner records could easily exist. To process such big datasets, however, becomes almost impossible with the existing storage techniques and query writing tools in governments. The national system should therefore not be a replication of the operational unit record system of the institutional and subnational level. The only learner-unit level records at national level should be records in the national learner identifier system for learner registration and learner tracking purposes.		
Directory	The system has the feature to **manage a directory of individual learners**.	
	The system has the feature to **add and delete individual learners**.	**Caution:** Ensure that when a learner is deleted that the historical data are not deleted. It may be best to *close* a learner's enrollment in a certain school rather than delete the learner, to keep the historical data intact.
	The system has the capability to **import learner information** from other systems (e.g. Excel, csv).	
Unique Identifier	Allocate a unique identifier for an **individual** according to prescribed algorithms.	**Note:** A unique identification code must be assigned to every learner. It is important that this identifier is consistent and accurate over time. A unique identifier is a single, nonduplicated number that is assigned to, and remains with, a learner throughout his or her education career irrespective of whether the learner changes schools. **Tip:** The national office should assign each learner an unique national learner identifier (registration number). The national office develops procedures to ensure that two identifiers of the same type are not assigned to the same learner. **Note:** The system uses a specific number of data items about the characteristics of a learner. These data items are used in an algorithm to allocate a unique identifier to a learner.
Questionnaire Design	The system has the feature to **set up screens** according to the questionnaire.	
	The system has the ability to add fields to the unit-level questionnaire. It is **customizable and flexible** to add fields according to the use case.	
Baseline Data Transfer	The system has the capability that enables users to **transfer existing data** from the previous year into a **new database** for all learners (roll over).	**Tip:** *Longitudinal data coverage* – The learner's unique identifier makes it possible to follow a learner's progress in the system through the identifier in longitudinal data. (Longitudinal data are data gathered on the same learner from year to year.)
Data Entry	The system has the capability to set up **data entry at any level** (national, subnational and institutional levels).	
Data Entry: Technology	Data entry takes place on a **computer keyboard locally** as a manual process.	**Note:** This functionality should exist on a local computer without connectivity.
	Data entry takes place over a **browser** using internet protocols or automated via an API.	**Note:** This is especially advantageous in a decentralized EMIS. **Caution:** Ensure that the bandwidth and connectivity adhere to system requirements.
	The system has the feature to enable organizations to develop and deliver content to **mobile devices** and takes advantage of mobile devices' native capabilities.	
Data Entry: Quality Assurance	The system has **built-in quality assurance rules**, such as authentication and validations.	**Note:** The system allows the user to define various validation rules to identify violations. At the point of data entry, a check should be made to see if the data falls within acceptable range levels of minimum and maximum values for any particular data element. As noted elsewhere in this guide, this may be easier to implement through background ratios (e.g. a pupil-teacher ratio that can flag if the enrollment data are not consistent with the teacher data). **Caution:** One must make sure that the built-in validation and verification rules do not slow down the data entry process. Furthermore, the validation rules should alert the user but must not stop the user from continuing with the data entry process.

DATA COLLECTION LAYER (GENERAL) (applies to either aggregate or unit-level data)		
Feature	**Description**	**Essential points to consider**
System Type Choices	**Operational transactional processing (OTP) system:** An OTP system is a system to capture and process the detailed (individual) information necessary to update data on the fundamental operations of an organization. A system used for the day-to-day running of the school.	**Note:** *Advantages*: • Real-time data available. • Data collection is not an extra burden because it is part of the day-to-day running of the school. • Increases the quality of the data. *Disadvantages*: • Security concerns. • Expensive to set up.
	Learner registration tracking system: A system to store the unit-level record data of learners and teachers in a central national database with the function to track the movement of individual learners from school to school and from year to year.	**Note:** The system is not a school management and administration system (operational) in the true sense of the word. It is mainly used as a learner registration and tracking system. The system is centrally developed and managed by the national office. The main functionality of the system is to register learners and to track and monitor individual learners in the country via registration of learners, transfers of learners between schools, etc. *Advantages*: • Easier to track the movement of learners from school to school and from year to year throughout their school careers. • This makes it possible to create a longitudinal dataset and determine exactly how many learners of a specific cohort dropped out of the system, how many progressed through the system without any repetition and how many are still in the system with one or more repetitions. *Disadvantages*: • Security concerns because of the personal details of individuals. • The volume of the data can be problematic with traditional storage methods. A further challenge is the processing speed to process and manage such large volumes of data.
	Aggregated data collection system: The system collects summary (aggregated) data from institutions for reporting and statistical purposes. Data collection takes place by means of an **annual census questionnaire** (the school census approach) sent to schools to collect the education information that the government needs to monitor the education system.	**Note:** Aggregated or summary data collection provides information on issues at institutional level, such as enrollment by grade and age, and enrollment by gender. **Note:** *Advantages*: • No privacy concerns because it does not contain personal identification. • Easier to identify patterns and trends in the data. *Disadvantages*: • Data initially collected at the institutional level cannot be used for lower levels of aggregation, for example, the tracking of individual students over time and across institutions. • Data quality is always a concern.

DATA COLLECTION LAYER (GENERAL) (applies to either aggregate or unit level-data)		
Feature	**Description**	**Essential points to consider**
Application Software Choice	The system is **custom made** and self-developed for a specific purpose (one of a kind) and owned by the organization.	**Note:** In the past, the only option that was available to ministries and governments was to develop EMIS solutions from scratch, which were either developed in-house or custom built by a software vendor.
	It is an **off-the-shelf** (mass-produced) system.	**Note:** The system is a management information package, which is available on the market and can fulfill many of the tasks needed within an EMIS. Such systems often require only minor modifications to the off-the-shelf product. **Note:** There is a distinction between commercial off-the-shelf (COTS) and government off-the-shelf (GOTS). The latter often has different acquisition requirements or expectations and will certainly have a different upgrade life cycle.
Operating System	The operating system on which the EMIS backbone runs serves as a platform by providing an environment that supports **user interaction through interfaces** (e.g. monitor, mouse, keyboard, printer). For example: Windows, macOS, Linux, Android, iOS and Chrome OS.	**Note:** *Ease of use, visual appeal and workflow integration:* refers to the ease of administering and deploying the platform, creating content, and consuming and interacting with content, as well as the degree to which the product is visually appealing.
Training	The system provides **training support**; i.e. it has good "help" functions at a minimum, linked training tips or, even better, manuals.	**Note:** Management should implement processes to ensure that the education ministry has an appropriately deployed EMIS workforce with the skills necessary to achieve EMIS goals. **Note:** The system should have a complete set of online Help search menus, FAQs, etc. as well as a well-indexed stand-alone guide. **Tip:** Ensure that EMIS staff has basic database skills and the ability to write simple queries to extract data from the database system. Queries can be automated, but ensuring that (at least some) EMIS staff can write queries (or even code) is a good way to set a bar on skills and ensure flexibility rather than be totally dependent on automation.

DATA MANAGEMENT AND DATA STORAGE LAYER		
Feature	**Description**	**Essential points to consider**
Database Types Choice	**Flat file system:** The system operates as a flat file (no relationship between records).	**Note:** The system uses a simple structure (single table of data) and cannot contain multiple tables compared to a structured database with a relational model. **Caution:** Avoid using this kind of database structure. A flat file system becomes increasingly inefficient as more data are added because there are no relationships between different sets of data. Though data are easier to understand and visualize in a flat file structure, accessing and searching through the data can become extremely slow and cumbersome.
	Single-user system: The database on the system installed is meant for a single user (stand-alone application). Stand-alone applications are traditional software that are installed on each user's computer.	
	Multiuser system: The database management system is meant to share information throughout the organization over a network. Many users can access the database at the same time. Naturally this is the preferred option in almost all cases. Some legacy systems may not have this functionality.	
Database Management	The database management system (DBMS) has a **relational** structure.	**Note:** A relational database management system (RDBMS) is a database management system that allows the temporary or permanent joining of data tables based on a common field (a primary and foreign key). Each row, record or instance in a database has a fixed set of attributes or fields. Each table has a primary key that uniquely identifies each record. The table may also contain a foreign key, which is identical to a primary key in an external table. A relational join is achieved by matching the values of the foreign key to the corresponding values in the primary key of the external table. An example would be a "table" of schools such that (i) for each school in that table there is a link to all the classrooms in that school, and (ii) those classrooms themselves are tables of data characterizing the classroom, without having to have the classrooms show up as nested within the school in the same table as the schools.
Software Types Choice	The system is developed with **open-source software**.	**Note:** Open-source software is distributed for free (depending on the use case) with the source code accessible and available to the users. *Advantage*: Very low cost and much greater flexibility in the infrastructure. **Caution:** The software cost is a small part of the total cost. Software support, software maintenance and in-house technical skills are key requirements and important. An open-source software solution often demands a greater technical understanding from the user, which could inadvertently increase costs.
	The system is built with **propriety software**.	**Note:** Propriety software is owned by an individual or a company. Because the software solution is owned, the users generally do not receive access to the source code. A key benefit is that propriety software comes with guarantees and support that open source usually does not provide.
Web-Based System	The database application is designed to be managed and accessed **through the internet**.	**Note:** A web-based system also offers the option to be deployed over a local or wide area network. **Tip:** Ensure that the system has an offline feature, which will enable data entry without internet connectivity. This means that if the internet goes down during a working session, the user can continue to enter data into the system and then upload the data once connectivity is restored.
Data Storage	The system has the capability to **store data from multiple years** in the same database.	**Note:** When large amounts of data for multiple years can be stored in the database, a longitudinal dataset can be created. Longitudinal data coverage is a key requirement in order to track individual learners through the education system. It also allows for easier tracking and portrayal of trends at the school level, a key value-add analysis feature that the EMIS ought to enable. This involves creating a dataset that includes information of the same learners from year to year.
Data Warehousing	The system uses a **data warehouse** or **data mart** for data mining.	**Note:** Transactional data are extracted from operational systems and transformed into one or more databases in the data warehouse for data mining and analysis using data warehouse protocols and principles. For more detail on data warehousing, see Section 6.2.5.3.

DATA MANAGEMENT AND DATA STORAGE LAYER		
Feature	**Description**	**Essential points to consider**
Security and Confidentiality	Strict access control, data security, privacy, data protection, confidentiality, authentication and encryption are key features in the software solution.	**Note:** To maintain security, capabilities are required that enable platform security, such as encrypted transfer of data (e.g. HTTPS, HSTS) and encryption-at-rest; administering of users; and auditing of platform access and utilization. A flexible security architecture should exist to manage access to the system at different levels by different types of users and custom roles. These are defined based on a country's requirements.
	Authentication: Users are uniquely identified and verified in the system.	**Note:** Assigns a unique identifier to all users in the system that is used to authenticate the user into the system.
	Password control: Every user may have a unique username and password linked to individual or group roles and privileges.	**Note:** The system has a feature to assign a unique username and password to each individual in the system. Each unique user is assigned a role within a group. **Note:** Different password types (see below on "hierarchy of access") can give the user different rights (e.g. read-only, read-and-write).
	Hierarchy of access: System ensures that users have access only to the data they require.	**Note:** This is achieved through a combination of users, roles, groups and privileges. "Roles" refer to features users have access to, and "groups" refer to the scope of accessible data. Users are created in the system with access to a certain level of the hierarchy where they work. The user can capture or access data only linked to this hierarchy based on the user roles and user groups.
Interoperability and Data Integration	The system is **compatible** with other existing systems.	**Note:** Interoperability refers to the communication protocol between systems. It is the ability for various information systems to exchange, share and use data cohesively. **Note:** Integration is the process of creating or ensuring interoperability. Interoperability is the process of combining or integrating heterogeneous data residing in different sources and providing the user with a unified view of these data.
	The functionality of being able to **export and import data from other sources** exists within the system.	
	Data from multiple sources can be **linked, integrated or merged** using a common field across a collection of data sources.	
Imputation	A good information system has the ability to impute data when the **data for particular schools are missing**.	**Note:** Appropriate editing and imputation methods are used and regularly reviewed, revised or updated as required. The imputation method used can determine how accurate the imputed value is. Information (detail) should be provided on why the particular method(s) was chosen and when it was last reviewed.
Web Hosting	**Cloud hosting:** Access to software and data on a server over the internet that is owned and managed by a third-party organization.	**Note:** Access to software on the cloud may be less controversial than having, say, student data stored on the cloud. *Advantage*: Access to data is possible from anywhere where there is internet connection. With cloud storage any change to the data is updated across all devices that have cloud access. *Disadvantage*: Cloud storage is dependent on internet connection. Privacy and security could be compromised and be more difficult to manage because the data is no longer on a physical storage environment. **Caution:** Ensure that cloud servers are reliable and highly secure, and all data are encrypted. A proper enterprise cloud provider should be able to show the appropriate security documentation or certification.
	Self-hosting: Installs and accesses software from their own server. Self-hosting is also often referred to as "on premises" or "local" hosting.	
	Software as a service (SaaS): A software licensing and delivery model in which software is licensed on a subscription basis and is centrally hosted. It is sometimes referred to as "on-demand software," and was formerly referred to as "software plus services" by Microsoft. SaaS applications are also known as web-based software, on-demand software and hosted software. Payment can take place by paying a monthly service charge or a per-use fee.	

DATA REPORTING AND ANALYSIS LAYER		
Feature	**Description**	**Essential points to consider**
Online Analytical Processing (OLAP)	A web portal for online analytical processing exists. The system allows users to **analyze information from multiple database systems at the same time**.	**Note:** A web portal is a website that functions as a single point of access to information from the national data warehouse and other relevant sources. The web portal's role is to make such data sources easily accessible in a structured, systematic way. The portal can connect to the data warehouse web interface and communicate with relevant resources such as maps, charts, reports, tables and static documents.
	OLAP has the functionality to **drill down** to the lowest level of the data.	
	OLAP has the functionality to **slice and dice**.	**Note:** OLAP is a computer process that enables users to select and extract data from different viewpoints. Slice and dice means to divide data into smaller parts (e.g. break down by school and then teachers in the schools and then the pupils of those teachers) to examine data from different viewpoints so the user can gain further insights.
	OLAP has **pivot table functionality**; i.e. the ability to create tabular data from a true database. In a true database each row might state the gender of the student, for instance. A pivot facility would allow the user to create a summary table where gender is a column or row.	
Dashboard	The system has the capability to create **highly interactive dashboards and content**, and visually tracks, analyzes and displays data. Data are visualized on a dashboard as tables, line charts, bar charts, etc. The visualizations on a dashboard can come from a single dataset or from many underlying datasets.	
Query Writing and Reporting	**Operational reporting:** Reports run on a scheduled basis and are directly distributed to the relevant users.	
	Self-service reporting: The user executes these reports from a simple menu-driven interface.	
	Parametric reporting: User-controlled parameters help the user to tailor the report to specific requirements at the time of execution. Parameters allow the user to alter the content and provide flexibility with limited user effort or technical knowledge.	
	Ad hoc queries: Ad hoc queries are written by the user and then stored for future use.	

5.5 Additional Minimum Requirements and Standards Related to Technology

Building on the previous section, some minimum requirements for the technology that underpins an EMIS are further elaborated here[14] – "technology" referring to *hardware*, *software*, *databases and networking*, and *telecommunications and networking*. In countries with very low income per capita or low overall IT infrastructure in government, these requirements may be overambitious and instead signal a growth path. For others, current technology in use may already have surpassed what is listed here.

Hardware consists of computer and related information and communication equipment used to perform input, processing and output activities. The hardware is the most visible part that plays a critical role in the functionality and success of the information system, but hardware features change constantly and there is too large a variety of hardware options that could underpin the software

14 There is therefore some repetition of key concepts from the previous section, as considered necessary.

requirements listed below. Thus, we focus on minimum requirements for the remaining technology components are discussed in the rest of this section.

5.5.1 Software requirements

Software is the set of instructions that tells the hardware what to do, consisting of two main categories: **operating system software** and **application software**. In order to navigate the vast number of software components, decisions need to be made first about the modules needed: the system architecture and the level of complexity. The following are the main components of a sound system software.

Operating system software: As discussed in Section 5.4, the operating system serves as a platform by providing an environment that supports user interaction through interfaces (e.g. the monitor, mouse, keyboard, printer). Examples of operating systems include Windows, macOS, Linux, Android, iOS, and Chrome OS. The operating system ought to make the EMIS easy to use and visually appealing, and allow workflow integration.

Application software: In many countries there may be legacy systems, and "deciding" on software for a whole EMIS may not be required. Nonetheless, decisions may still need to be made regarding other systems that interact, manually or not, with the legacy system.

Custom-made. The advantage of custom-made software is that it can be developed according to the specific needs of the country. Because the ministry is in control of the process and the end result, unnecessary features can be avoided. However, to custom build software can be a long and slow process, and the initial cost very high.

Off the shelf. Such operating systems often require only minor modifications to the off-the-shelf product. Off-the-shelf software can have a number of benefits/advantages, including:

- The modification and implementation of the software could be a faster process. However, depending on the use case, modification of commercial off-the-shelf (see previous section), software can sometimes be difficult or questionable.
- The software solution could be a cost-saving exercise.
- With off-the-shelf software, features can sometimes be tested before buying it.

The disadvantage of off-the-shelf software is that it often comes with unnecessary features that may add unnecessary costs.

Application software: desirable features

Front end: Whether data are entered electronically at the local level or manually at the central level, the software needs to have a data entry module with assigned functionalities and a built-in validation process. This is often referred to as "the front end".

Data validation functions: Data validation must be built into the different modules at all stages of the data production chain: data collection, entry and dissemination.

Application software: types

Open-source or free software[15] This is distributed free of charge with the source code accessi-

15 Take care to distinguish between "free" and "open source." There are many "free/freemium" proprietary applications, especially in the education space, while not all open-source applications are free all the time. Furthermore, within open source are "copyleft" licenses, which may preclude the ability of certain commercial partners to

ble and available to the users. As detailed earlier, key benefits are the very low cost and much greater flexibility in the infrastructure. However, the software cost is a small part of the total cost. Software support, software maintenance and in-house technical skills are important. An open-source software solution demands a greater technical understanding from the user, which could inadvertently increase costs.

Proprietary software. Because the software solution is owned – whether by an individual or a company – the users generally do not receive access to the source code. A key benefit is that propriety software may come with guarantees and support that open source usually does not have. Note that if the ministry chooses to develop its own custom-made software, then it could claim intellectual property over it, depending on the country's rules.

5.5.2 Database requirements

Database approach and databases

Relational database management system (RDBMS). This is the most frequently used database management system (DBMS),[16] and allows the temporary or permanent joining of data tables based on a common field (a primary and foreign key). Each row, record or instance in a database has a fixed set of attributes or fields. Each table has a primary key that uniquely identifies each record. The table may also contain a foreign key, which is identical to a primary key in an external table. A relational join is achieved by matching the values of the foreign key to the corresponding values in the primary key of the external table.

Flat file system. As detailed earlier, the system uses a simple structure (a single table of data) and cannot contain multiple tables compared to a structured database with a relational model. This kind of database structure should be avoided because, given that there is no relationship between the data, a flat file system becomes increasingly inefficient as more data are added. Though data are easier to understand and visualize in a flat file structure, accessing and searching through the data can be slow and cumbersome.

Data handling via spreadsheets. If spreadsheets have been used in the past, systems should quickly migrate away from them. When used for a task they are not designed to perform, such as a database or statistical software, or for a task beyond their capabilities, spreadsheets can be risky. As information is formatted in the actual spreadsheet, it is cumbersome to retrieve data from the simplest queries. Although a spreadsheet allows for the use of worksheets to create relations between tables, such relations will be very limited. Furthermore, the bigger a spreadsheet becomes, the more memory is needed to keep it open and the greater the danger of major mistakes.

NoSQL (not only SQL) database. This provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases. This is a nonrelational database of the sort used to track event streams.

Data storage. The system has the capability to store data across multiple years, thus enabling the seamless transfer of the information from one year

take part in an open-source activity, as well as licenses that are more appropriate for commercial open-source development.

16 A DBMS is a software package that allows users to create, retrieve, update and manage data.

to the next. If there is no data warehouse available, ensure that the database is designed in such a way for EMIS to store data from current and previous years. Historical data are important to analyze trends. Establish a data retention policy, because some data should be retained for many years, while other data may only be needed for days.

Imputation. It is important that appropriate editing and imputation methods are used and regularly reviewed, revised or updated as required. Imputation is a method for estimating missing values. The simplest example would be a school that has missing data for a particular variable (e.g. number of books provided to the school in given year) but has data for years past, in which case an extrapolation could be made. Or, for example, if it is known that books' provision is based on a formula driven by enrollment, the formula could be used to impute the values. As detailed earlier, a good information system should have the ability to impute data when the data for particular schools are missing. The imputation method used can determine how accurate the imputed value is. Information (detail) should be provided on why the particular method(s) was chosen and when it was last reviewed.

Web-based system. To the extent possible, the database application ought to be designed to be managed and accessed through the internet. An internet-based server architecture using browsers and APIs may be more easily deployed (use case dependent) than a traditional client server system based on a structured set of interactions between a computer acting as file server and users with particular needs for accessing those files. As detailed earlier, a web-based system can optionally be deployed easily over a local or wide area network or the internet. Ensure that the system has an offline feature, which enables data entry without internet connectivity. This means that if the internet goes down during a working session, the user can continue to enter data into the system and then upload the data (automatically, ideally) once the connectivity is restored.

Supporting tools

Definitions of education terms (dictionary of terms and concepts). This is a prerequisite for the effective functioning of any information system as it provides a standard for the terms and concepts used by an organization in its information collection and reporting processes. It often happens that in information systems different terminologies are used to describe the same concept or principle in different publications and databases. In a good system, terminology should be consistent across institutions so that ambiguity does not arise. The dictionary could be part of the help system of the software platform but should also exist apart from the platform.

Data dictionary (metadata) and meta-tagging system. The former (a centralized repository of information *about* the data) explains to the user what the codes in the data tables mean (e.g. that the ages of students may range from x to y, or the range and descriptions of valid answers to, say, the "principal language of instruction") mean. The latter provides a limited or recommended vocabulary for describing data and/or content in a system (that is, in a sense, instructions on how to create the data dictionary).

5.5.3 Telecommunications and data networking requirements (including networks and the internet)

Communications and being connected are key characteristics of modern social information systems. A good EMIS design should consider the following:

Bandwidth. Bandwidth is a measure of how much data can travel on the network at one time, at various points in the network. Effective bandwidth is important for internet connectivity and refers to the maximum and best reliable transmission rate that a network path can produce. In some networking platforms like home broadband, the effective bandwidth is limited. Bandwidth is highly significant for determining how fast a web page loads onto a browser. The bandwidth can be significantly affected by the website and the internet connection used for accessing it. If storing data on servers in the cloud is the main approach used then bandwidth is arguably less important; if one is transferring data from one location to another, or entering remotely, then adequate bandwidth is crucial.

Mobile technology. A mobile service (smartphone and tablets) is intended to enable rapid mobile data collection using both online and offline modes. Note that mobile phone coverage must be taken into account.

Technology readiness: This is a key aspect for the implementation of mobile technology for data collection, and considerations include device ownership (including price per unit, warranty and repair), internet connectivity and data costs. Technology with offline data collection may be a strong option for low-income countries and rural areas.

Baseline data transfer. The system ought to enable users to transfer existing data from previous years into a new database for all institutions. As detailed earlier, this ability is key to the generation of time trends and other forms of value-add data. Where applicable, the system should have the option to preprint (or pre-populate if online) baseline data on the physical questionnaire before it is disseminated to the institutions. Note, however, that pre printing (as opposed to online pre-populating) vastly increases the logistical requirements as exactly the right questionnaire has to get to exactly the right school.

5.6 Future Technology Considerations

Next generation conceptual frameworks for EMIS design now have begun to focus on the vast potential of data that are collected primarily from schools, students and staff to improve learning outcomes. Realizing this potential within the context of EMIS architecture requires a significant rethinking and restructuring of data systems within the education sector. This envisioned restructuring of EMIS platforms would place data centrally in policy decision-making processes and vastly improve the flow of data to ensure that it can be used reliably by a wide variety of stakeholders: administrators, teachers, students, guardians, employers.

Modern EMIS technology can tap the vast amounts of education data that are collected and lying dormant. Typically, much more data are collected than analyzed to yield insights for education stakeholders. Legacy EMIS architectures are often unable to adequately address barriers to timely data use, such as misaligned data capture resulting in data gaps, disorganized and incompatible subsystems, latency in reporting, and a fundamental lack of trust in the quality of the data. Modern EMIS technology can help overcome these challenges.

Legacy RDBMS/OLAP technology can now be augmented with new approaches to data ingestion, transformation, storage, analytics and visualization. New EMIS data architecture can

provide scalable and efficient data pipelines with high throughput and low latency.[17] A well-designed EMIS data pipeline can deliver an end-to-end platform for collecting education data from various sources, transforming the data into insights, and delivering the information to end users in a timely manner. Table 3 attempts to summarize these future-oriented issues, partially by comparing to the current issues. Note that the table cannot possible be exhaustive; only a few important examples are shown.

17 *Latency* is the time taken for a packet to be transferred across a network. *Throughput* is the quantity of data being sent and received within a unit of time.

Table 3. Technology Upgrade Paths

Layer	Currently typically handled by[18]	Future or even current possibility to consider
Data Collection Layer	Paper, Excel, Access	Progressive web apps are becoming increasingly powerful and user-friendly.
	Paper, web-based, but still keyboard based	For bulk-type operations, scanning of QR codes from phones or tablets. Biometrics also increasingly available (scan of fingerprints, whole hands, iris).
Data Management and Storage Layer	Local files, local databases, email attachments	Extensive use of online resources with offline alternatives for poor connections (paper, SMS). Authoritative services for key metadata (master school list, human resources, student register) and data that can be stored as a combination of centralized data in the (national) cloud and data retained locally.
	Data are "intentional" – gathered as data, typically for administrative or policy/ planning purposes.	Data could be "residual" or by-products of social media scraping, complaints forms (even telephone), etc.
Data Analysis and Reporting Layer	Desktop (single user) tools, mainly Excel, but also specialized analytical tools, such as Stata, SPSS, ArcGIS. These often require significant pre-processing of data, as one-off events performed by experts.	Integrated apps for interactive analysis become increasingly available; e.g. R Shiny and a host of JavaScript visualization tools can access data through APIs and be made accessible to a large number of users over the internet. User interfaces can be tailored to specific needs, and analysis can be shared online, through dashboards, and interpretations discussed via forums and chats. Simulations and machine learning tools are also increasingly available, and especially useful for planning purposes.

Source: Authors.

18 This column is somewhat weighted toward low-income countries, not lower-middle-income countries, and certainly not upper-middle- or high-income countries.

6. User's Guide

Once officials and supporting agencies have determined the functionalities they wish to have in an EMIS, studying how that system can be put to use would be the next step. This section elaborates on this task. In addition, the UIS has prepared the *Operational Guide to Using EMIS to Monitor SDG 4*, which delves into more specific terms on how an EMIS could be used to report on SDG 4, which can be found at: http://uis.unesco.org/sites/default/files/documents/operational_guide_to_using_emis.pdf

6.1 Staff Requirements

In addition to being able to understand specific functionality standards, and then show how to apply them, a user's guide needs to help users determine whether staff have the skills needed to run the systems. Box 4 shows a snapshot of some of these skills, at a fairly high level.

6.2 EMIS Production Life Cycle

The User's Guide sets out processes, procedures and practices for collecting, processing and disseminating education data to ensure the production of quality data – namely, data that is *complete*, *relevant*, *accurate*, *timely* and *accessible*.

For the purpose of this guide, these processes are referred to as the *EMIS production life cycle*,

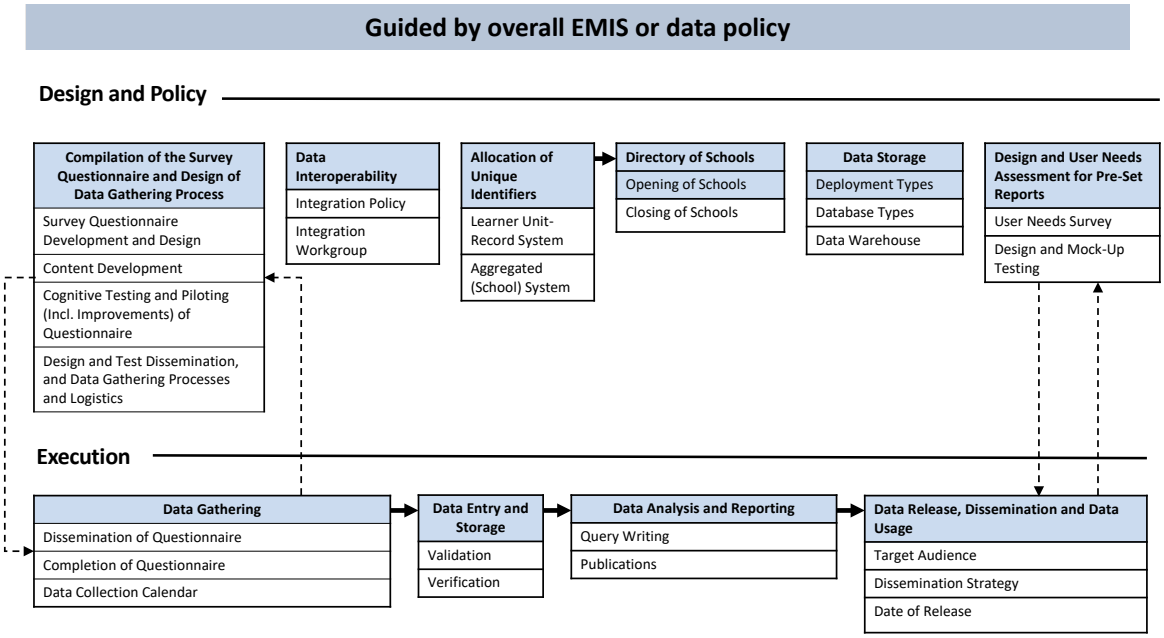
Box 4. Technical Skills for EMIS Staff at National, Subnational and School Levels

The organizational structure of EMIS should be an expression of the dominant work-related processes and activities. The success of an EMIS initiative in the MoE rests on how effectively it is able to leverage the skills and knowledge embedded within its staff. It is important that staff with the relevant skills profile be present at national, subnational and institutional levels in EMIS, and that these reflect the MoE's information needs and strategic direction.

The skills profile of EMIS staff at all levels should include at least the following:

- Knowledge of data collection processes and activities.
- Basic database skills (database essentials including complex query writing).
- Knowledge in the management and analysis of data (including, if possible, at least a basic course in statistics).
- Basic ICT skills and knowledge, where ICT refers to several different components that enable users/staff to retrieve, store, produce and share information, including:
  - Hardware – computers, printers, etc.
  - Software – operating systems and programs (e.g. spreadsheets, databases, presentations) installed on computers.
  - Communications systems – internet, browsers, email, cell phone information gathering, etc.
  - Online platforms – video-calling (e.g. Skype, Zoom, Hangout, Teams).

Figure 3. EMIS Production Life Cycle



Source: Authors.

Note: While the production process in the bottom panel – execution – is relatively linear, the top panel – design and policy – is relatively nonlinear,[19] with only a few of the steps being sequential. For example, it is difficult to have a proper directory of schools unless unique school IDs have been assigned. Not all the areas that are linearly related in the top panel are depicted, so as to keep the diagram uncluttered. Just one of the arrows is drawn in as illustration. There is interaction between the two panels, as design is of course necessary for execution and there are feedback loops from execution to design; indeed, there are in reality far more feedback loops than are depicted here. Note also that "usage" refers more to an everyday or managerial usage than usage for policy and planning. Usage in the latter sense would happen outside of the life cycle depicted here, in other units of the ministry (see Section 4 on data architecture).

19 That is, more linear than the top panel, and more linear than the interaction between the top and bottom panels.

as depicted in Figure 3. The EMIS production life cycle directs the reader step by step through these processes and practices, from the initial design and development of the survey questionnaire through to the data collection processes to the use and dissemination of the data. The User's Guide will rely on the EMIS production life cycle as a lens to determine all the key activities in the data collection and usage processes at all levels, from national to subnational to schools.

6.2.1 Compilation of the survey questionnaire (design and development)

At the heart of the EMIS lies the survey questionnaire. In most developing countries, the data required for the EMIS are obtained through surveys (the school census approach – see Box 5). Data collection can be manual-based (paper-based) or technology-based. The data collection from schools, the primary data source, via a survey is usually done with a questionnaire that also serves as an instrument to develop the database. In a unitary system, the national office (usually the EMIS unit at head office) is responsible for the compilation, design, printing (where applicable) and on-time dissemination of the survey questionnaire. In a federal system, the national level might convene agreements and collaboratively set standards, but typically the first subnational level

Box 5. School Census Approach: Manual or Web?

There are certain disadvantages to manual data entry of the school census. It is often time-consuming to print and distribute the survey form to all the relevant entities, and it can be extremely costly in terms of fuel and paper. It is logistically challenging, and some schools are almost always missed. The capturing and cleaning of the data also tend to be time-consuming processes. Consequently, quality and reliable data take too long to be available for education planning and decision-making purposes. However, there are some advantages to a manual system, especially in very poor areas, where not all schools have internet connections, or devices like tablet or PCs, or, if they do, are even reasonably able to upload asynchronously (that is, after the data have been entered on the local PC or tablet, they are uploaded to the server later, when convenient, such as Wi-Fi becoming available). Alternatively, one can have manual entries in some areas of a country and electronic in others, or manual at some levels (e.g. school) and electronic at other levels where there may be internet access (e.g. district offices).

(province or state) would run the operations, and standards may vary.

6.2.1.1 The design of the questionnaire

The initial design of the questionnaire is important for these reasons:

- Through the questionnaire, the input data from the schools are obtained, either through a physical form (paper-based) or via the internet in a web-based system.
- It is the most important building block of the database and used as the User Requirement Specifications (URS)[20] to develop the database model.
- It is used to design and develop the input screens in case of an electronic or web-based system (Box 6 describes a particular issue noted in previous applications that have both paper and electronic entry). Care should be taken that when the entry is not solely electronic, the electronic and paper-based versions both map in exactly the same manner to the questionnaire.

20 A URS is a document that the developers of the database use as a guide to determine which functions need to be included in the system and to create the data model.

Box 6. Discrepancies Between Manual Data Collection and EMIS Input Screens[21]

In the case of electronic entry, the input screen of electronic or web-based systems is an "exact" mapping of the content of the questionnaire. For paper-based surveys, the database input screens developed for data capturing may not represent the whole content of the questionnaire, which may lead to a "deliberate" gap between the data collected using the school census questionnaires and the data captured in the EMIS database. This uncommon practice was encountered in some countries in Sub-Saharan Africa during the EFA and MDGs period (due to limited system capacity or resources constraints).

21 Insight courtesy of O. Labé (see Annex A for a list of all contributors).

6.2.1.2 Content development of the survey questionnaire

- The content development of the survey questionnaire is a collaborative process; i.e. the content (questions) of the survey questionnaire is compiled in collaboration with key stakeholders.
- The content of the survey questionnaire is essential in ensuring that all relevant information will be captured.
- The content of the survey questionnaire consists of categories such as general information of the school (name, address, sector, grades) and learner (enrollment, home language, age and subjects offered).
- Repeated cognitive testing and piloting of the questionnaire need to be part of the process, even after an EMIS questionnaire has been in use, if cognitive testing was before (so as to improve it over time).

General principles that apply to the construction of good questionnaires are not elaborated in detail in this guide. For further guidance on the content of the survey questionnaire, it may be useful to look at questionnaire design as carried out by researchers working on household or facilities surveys, as they sometimes pilot test questionnaires repeatedly, sometimes use cognitive laboratories to understand how individuals interpret questions, and so on. There These practices may be worth imitating, as these researchers often devote more time to understanding how respondents react to questionnaires than the designers of administrative forms. There are also several institutions with resources on both random sample surveys of households and institutions as well as administrative form design.[22] While most EMIS in the lower-income regions are unlikely to implement direct entry by schools into computers or tablets in the immediate future, it is interesting to note that similar considerations apply to designing good online forms as to traditional paper and pencil forms.[23]

6.2.1.3 The dissemination of the survey questionnaire

- In case of a manual process, the survey questionnaire should be disseminated to all the relevant institutions for completion by using the directory of schools.
- In case of a technology-based process, access to the system should be provided to all the institutions on a timely basis.
- Both methods could co-exist in a country depending on the level of technology and the level of the system (school versus district, for example).

6.2.1.4 The completion of the survey questionnaire

- All institutions (e.g. primary and secondary; public or private) should complete the survey questionnaire before a specific date.
- The data elements on the survey questionnaire are considered imperative in the provision and management of education in the country.
- Excessively long questionnaires should be discouraged and details that may not be important or relevant should be avoided.

22 See, for instance: https://www.povertyactionlab.org/sites/default/files/documents/Instrument%20Design_Diva_final.pdf; https://github.com/worldbank/DIME-Resources/blob/master/survey-instruments.pdf; http://documents.worldbank.org/curated/en/452741468778781879/Volume-One; http://surveys.worldbank.org/sites/default/files/files/C4D2-T/C4D2-T2/C4D2-T2_PPTS/designer.pdf; https://unstats.un.org/unsd/EconStatKB/KnowledgebaseArticle10364.aspx. For examples of government guidelines on web-based forms, see https://www.usability.gov/get-involved/blog/2008/04/usable-online-forms.html and https://www.usability.gov/how-to-and-tools/resources/publications/defensive-design-web-how-improve-error-messages-help-forms-and-other-crisis-points.html. For an example of a consultancy-based set of guidelines, see https://www.effortmark.co.uk/. See Dillman et al., 2014 (in Suggested readings).

23 This insight courtesy of D. Dillman, Washington State University.

- Suitable guidelines (data dictionary and metadata) on how to complete the questionnaire should accompany the questionnaire and, if necessary, training.
- Any technical and data terms and elements used in the questionnaire should be clearly defined on the questionnaire so that data providers can understand what information they should provide.
- The completion of the questionnaire at school level should take place by using official records. For example, where age is required, the class register with details of the date of birth should be consulted.
- With regard to verification of the survey questionnaire (quality proof of reported data), it is to
  - Verify information, such as enrollment totals against the legal documents used for re-cordkeeping by the school (e.g. admission register); and
  - Verify that tables with the same information on the questionnaire, such as enrollment by age, or by grade or by gender, all yield the same total enrollment. Data entry could error trap these sorts of issues.

6.2.1.5 Data collection calendar

- A data collection calendar that covers all the steps in the data collection process should be disseminated to all relevant role players. An example is included in Annex C.
- These steps should be indicated on the survey questionnaire.
- Important dates should be included on the calendar, such as
  - Dissemination of survey questionnaires;
  - Completion of survey questionnaires at school level;
  - Start of the data capturing;
  - Completion of the data capturing; and
  - Release of the data.

6.2.2 The Allocation of unique identifiers

Unique identifiers are nationally generated (or provincially generated but with a national protocol even in a federal country) numbers allocated to all persons/entities (e.g. learners – admission number; teaching and other staff members – salary number; educational institutions – EMIS number).

In creating a longitudinal data system, it is necessary to link the different datasets that have been collected for individual learners or individual schools for each year by using a common field across these datasets. To be able to do this, unique identification codes should be assigned to every learner (see the discussion of "learner unit record" in Table 2 of Section 5.4) or to every school. It is important that this identifier is consistent and accurate over time.

The national EMIS office (or perhaps the provincial office in a truly federal country) should develop and introduce the standards governing the generation of unique identifiers.

6.2.2.1 The learner unique identification system

Learner unique identification system where a unit-level record EMIS is implemented

Individual learner tracking may not be suitable in many contexts as they are technologically and administratively demanding. In such situations a careful analysis of costs and benefits should be done prior to embarking on such a project. In cases where a learner unique identification system is suitable, the following considerations apply.

A learner unique identifier is a single, nonduplicated number that is assigned to, and remains with, a learner throughout his or her education career irrespective of whether the learner changes

schools. No learner has more than one number, and no two learners have the same number.

The learner unique identifier makes it possible to follow learners' progress in the system through the identifier in longitudinal data (data gathered on the same learner from year to year). The national office (or possibly provincial in a federal country, then reported up) assigns each learner a unique national learner identifier (registration) that can be used to match records accurately across years.

The basic functioning of the learner unique identifier system involves the following:

- The national office develops procedures to ensure that no two equal identifiers of the same type are assigned to the same learner (e.g. when the learner moves, he/she keeps the same identifier) and that no two learners are assigned the same identifier. These rules must be followed by all levels.
- The national EMIS (or provincial in a federal country) assigns an identifier that will follow each learner throughout all the grades as he/she moves (transfers) across schools or subnational regions and leaves (de-registration) and reenters (re-registration) the country's public education system.
- The national EMIS
  - Enables subnational levels to allocate new identifiers for learners who do not have an existing identifier;
  - Issues or records a reported-up national identifier (if in a federal system) that will be used by subnational levels to report unit-level data to the national education department; and
  - Relies on a specific number of data items that provide characteristics about a learner. These data items are stored in a secure and confidential database at national level and are used to identify a learner uniquely. These items could include among others, the learner's name, surname, date of birth, gender, etc.

6.2.2.2 Unique identifier system for institutions (schools and other venues)

Unique identifier system for institutions (directory of institutions)

Note that these are typically more important than learner identifiers, as some countries may not be able to use (or may not prioritize) unique learner identifiers, while unique school identifiers and a master list are key to any well-functioning EMIS.

The Master File System for institutions is a key process to provide a unique identifier to an institution and should be managed at central level.

The purpose of such a system is to assign a unique identifier to every institution in the country. The institution unique identifier makes it possible to analyze institutional longitudinal data (data gathered on the same institution from year to year). One may also want the identifier for each school to automatically contain the placement of the school in the governance hierarchy. That is, the identifier may consist of codes for the levels higher than the schools, though this path of hierarchical codes is not strictly necessary, as long as the database has fields (e.g. district, province) that allow one to identify the school.

The basic functioning of the institutional unique identifier system involves the following:

- The national (or, typically, the first level of subnational office in a federal system) assigns each institution a unique national institutional

identifier (registration) that can be used to match records accurately across years.

- The national office develops procedures to ensure that two institutions are not assigned the same identifier.
- The national office in conjunction with the subnational EMIS develops procedures for the closing and opening of institutions.

6.2.3 The Maintenance of the directory (register) of schools

A directory of schools should exist, and be maintained and regularly updated (quarterly, per semester or annually). More specifically:

- A dedicated unit responsible for the directory of schools should exist at a subnational (or perhaps sub-provincial in a federal country) level that is responsible for opening and closing of all private and public institutions (noting that some countries have very weak records on private schools, especially in pre-primary).
- The purpose of such a unit is that during the registration of schools, a unique identifier (in conjunction with national office) is assigned to every institution and to maintain and update the official directory of institutions and to make sure that the same identifier is not issued to more than one school.
- The directory of institutions is an important step in the entire EMIS life cycle in order to facilitate the dissemination of questionnaires (in a paper-based system) or access to questionnaires (in a web-based system).
- The response rate of the data collection can accurately be determined with an up-to-date and complete directory of institutions.
- The directory of institutions should include the basic contact details of institutions, curriculum and language information, even spatial information, required by the information user.
- A directory of all the institutions should be available at the head office and in each of the subnational offices.
- All entities in the ministry that might collect data independently of EMIS are required (even retroactively, that is, to populate the identifiers into legacy databases of any importance) to use the master list and the EMIS identifiers. Should ministry entities wish to use other identifiers as well, for their own purposes, care should be taken to ensure no confusion and that both identifiers always accompany the data records and are inherited by all databases.

6.2.4 Data entering

- Quality assurance is essential to produce quality data (see Section 5.3), data that are complete, relevant, accurate, timely and accessible.
- Data entry must be verified and validated, particularly in the case of a manual data entry process.

6.2.4.1 Verification

- Verification ensures the prevention of errors occurring when data are captured.
- Verification entails checking that data have been entered correctly from the survey form into the system.
- The verification methods that could be used include the following:
  - Checking that the data entered from the survey questionnaire correspond with the data in the database (proof reading).
  - Using computer programs (queries) to compare the total of data tables that should yield the same totals, such as

enrollment by age versus enrollment by gender versus enrollment by grade, etc.

6.2.4.2 Validation

- Validation means checking that the data entered have meaning; for example, that data such as enrollment, age, etc. are within a range of permissible values.

6.2.5 Data storage

The database for data storage is a key component of the EMIS life cycle. A database is organized in a way that the data stored are easily accessed, depending on the deployment.

6.2.5.1 Deployment type

- *Offline deployment:* Offline deployment means that multiple stand-alone offline instances are installed for end users, typically at subnational level. Primarily officials at subnational level maintain the system.

- *Online deployment:* The decision to use a web-based server architecture is because internet-based server architecture is more easily deployed than a client server system. A web-based system can optionally be deployed easily over a local or wide area network or the internet. Finally, the web-based architecture serves to facilitate the provision of data on a timely basis.

6.2.5.2 Database type

- It is optimal to use a relational database management system (RDBMS) for data capturing. The choice as to whether there is a desire for a RDBMS or not should be in the user specs.
- The relational structure makes it easy to query the database and to integrate large datasets from multiple sources.
- Data integration generally means linking different data sources through a common field across a collection of data sources.
- To be able to do this, unique identifier codes must be assigned to the datasets that are used for the integration.

6.2.5.3 Data warehouse

- A data warehouse is a set of interconnected databases designed and developed primarily to be used for analysis. Typically, data are uploaded from various operational or transactional systems and contain different, linked, databases.
- The data warehouse is built to facilitate the use of management information and not for operational purposes.
- The data warehouse should ideally use the same unique IDs as the EMIS backbone.
- Whether the data warehouse should be part of EMIS or some other office (such as analysis and planning) is optional.
- The source for the data warehouse is the operational databases from learner enrollment (EMIS), learning assessments (examinations), student health, teachers, learning material (textbooks), physical facilities, etc.
- It is the data warehouse that ultimately allows significant value-add to the EMIS backbone for policy analysis and more complex planning purposes. Raw EMIS data are useful for management and more routine planning (e.g. school resourcing).
- Unlike the tables in the operational systems, the normal-form rules do not apply and any denormalization in the design that will facilitate the information gathering process is acceptable. For example, a child's name would

not be reproduced each time an event in that child's life is recorded; the child's ID is used, and the name could still be found if needed.[24]

6.2.6 Data interoperability

Information is of far greater value when it is integrated. The development of standard definitions and coding schemes across the education system will result in more timely integration of data across units. Data integration generally means linking different data sources through the use of a common field across a collection of data sources (see Box 7). To be able to do this, unique identification codes must be assigned to every level of the education system for which data is collected. The lack of commonly used unique identifiers that allow linking across data systems contributes to the unavailability of integrated information systems.

Interoperability allows for different systems (levels) to "talk to one another" and share data. Data and information sharing is greatly facilitated by a common set of standards and metadata. Such technical standards and guidelines should describe ways to achieve the interoperability of government departments' data. It creates the opportunity for government levels (national, regional and local) to join and share its datasets. This will enable the seamless flow of information across government departments and levels.

Interoperability is particularly important in terms of reporting on the SDG 4 indicators, as many of them call for data not present in the main traditional databases of an EMIS and could pertain to what this guide calls a "data warehouse" (see discussion in Section 5.1.2).

6.2.7 Data analysis and reporting

The datasets generated by EMIS (and in some cases the related warehouses) are often underutilized data sources. An EMIS should therefore maximize the produced datasets through the promotion of data analysis and research practices. The objective of an EMIS is not only to

24 See https://en.wikipedia.org/wiki/Database_normalization.

Box 7. Interoperability Strategy

Often multiple components within the education system collect and manage large databases and do not share them with each other. For example, data on student assessment (and/or examinations), data on teacher qualification and salary (payroll), data on enrollment (EMIS), and data on supplies of learning materials (textbooks) are stored in separate databases and not shared. If these datasets are not linked, the datasets will remain fragmented and exist in isolation of each other, preventing the kinds of useful value-add analyses that really ought to be at the heart of the EMIS mission and are needed to ensure the budgetary health of EMIS.

To make data available for this kind of powerful analysis, one can

- Develop and implement a data sharing and data integration policy; and
- Establish a formal working group or task team responsible for linking datasets such as enrollment, assessment, payroll, learning material and teachers using a common unique identifier.

collect, store and process information but also to help in education policymaking and planning, by providing relevant and accessible information. Therefore, apart from producing the usual statistical reports with tables and graphs, the EMIS should look at tools to discover trends and relationships in the datasets that are produced. Business intelligence tools support this kind of data analysis and provide this functionality.

One way to increase both the rationality of decision-making and the demand for data is for an EMIS policy (or education data policy, more generally) to enunciate the kinds and range of decisions that must be based on evidence shown. Additionally, EMIS units could encourage different positions held by think tanks and civil society to be argued or defended on the basis of data. A good example is whether dropping out or repetition and over-enrollment are the bigger problem in the early grades, or just how much dropping out there truly is in primary school (versus low learning or repetition or both, for example).

6.2.7.1 Query writing and reporting

The following types of reporting are recommended:

- *Operational reporting:* Reports run on a scheduled basis and are directly distributed to the relevant users.
- *Self-service reporting:* The user executes these reports from a simple menu-driven interface.
- *Parametric reporting:* User-controlled parameters help the user tailor the report to specific requirements at the time of execution. Parameters allow the user to alter the content and provide flexibility with limited user effort or technical knowledge.
- *Ad hoc queries:* Ad hoc queries are written by the user and then stored for future use.
- *Online analytical processing (OLAP):* Users can analyze and slice and dice data across multiple dimensions (local district, gender, wealth, etc.) to get a handle on apparent causes of issues such as low school completion. With this functionality, users can perform very complex calculations on large amounts of data. Alternatively, data can be downloaded and processed with tools as simple as Excel or as advanced as Stata or R.

6.2.7.2 Publications

This EMIS should optimize datasets through the promotion of data reporting, data querying, data analysis and research practices. The EMIS annual report on statistics on education should be disseminated widely within 12 months of data collection, according to these recommendations:

- EMIS data should be accessible at a subnational level.
- An EMIS report should also be published at a subnational level in the same format as the national EMIS report.
- Staff with query writing competencies and skills should be available within the EMIS function at national and subnational levels as one cannot rely on “canned” or ready-made “usual queries”; ad hoc queries are often necessary.
- Ensure that EMIS staff have the basic database skills and the ability to extract data from the database system.

Make the EMIS data available electronically (on the internet) in a user-friendly file format, such as a database, spreadsheet or even a comma-delimited text file that could be downloaded, while respecting strict confidentiality rules and “right to use” procedures.

6.2.8 Data release, data dissemination and data usage

The data that EMIS produces should be disseminated and shared with all the relevant stakeholders at all levels of the education system to support decision-making, policy analysis and formulation, planning, monitoring and management.[25]

6.2.8.1 Data release and data dissemination

As per the calendar discussed in Section 6.2.1.5, the EMIS should:

- Disseminate and make accessible (especially at subnational level), relevant, high-quality, timely and accessible statistics that will meet user needs in government.
- Data not routinely disseminated (microdata and alternative arrangements of published data) should be made available upon request, under a legal framework including provision for confidentiality.
- Since most schools are not interested in detailed data for the whole country, a paper dissemination strategy could consist simply of sending to each district a "canned" (e.g. PDF) three-pager for each school that has the key data (especially data that can compare inputs to outputs and produce some key ratios) for the school, its district context, and the national context. The district can then print 300 pages (3 pages for 100 schools – purely an example) and distribute. This may be especially useful where there is low internet connectivity or low capacity in the schools to run even simple queries.
- A data dissemination strategy should exist and should
  - Identify the information requirements of the target audience;
  - Determine the medium in which the different types of documents need to be produced (paper or electronic);
  - Propose the time period of the hard and e-copy data releases;
  - Identify the official or unit responsible for the data release; and
  - Determine what will be released (e.g. enrollment, repetition, any relevant issues).[26]

6.2.8.2 Data usage

Recent years have seen a considerable expansion in the availability, and in some cases also, the quality of data available for policy- and decision-making and research. This guide contains some suggestions regarding accessibility of data in a form that could make data better utilized for policy, analysis and planning. EMIS should actively promote the use of statistics and regularly consult with users.

Three recommendations are pertinent in this context:

- Provide ready-made "usual queries" that the user can adapt to their own needs (e.g. a particular school or district).
- Increase the accessibility of the annual school census data by making it available in a user-friendly file format, such as a database, spreadsheets or even a comma-delimited text file that could be downloaded.

25 For an example of a data release plan (from the U.K.), see https://www.gov.uk/government/organisations/department-for-education/about/statistics#ad-hoc-statistics-and-data-releases or https://tinyurl.com/y9cudbvm.

26 For some ideas and examples, see the World Bank's suggestions at: http://documents.worldbank.org/curated/en/155601468205458626/pdf/466170NWP0Box31nformation0System111.pdf or https://tinyurl.com/y9kmtapd.

- Do not provide users with more data than they need or than are interesting, as this may lose their attention.
- Allow the user at least basic choices over what to download in terms of records and record groupings (schools or school groupings, either singly or in aggregate, such as all schools in a district individually, or the total or average of all schools in a district), much as is possible in an Excel pivot table) as well as fields (variables), again much as in an Excel pivot table.
- Data dissemination and annual school census reports should be available online, which could increase the usage of data if the majority of users are connected to the internet.
- Make the EMIS data available with documentation, such as the user guide, codebooks and the data files.

Suggested Reading

Abdul-Hamid, H. (2017). *Data for Learning: Building a Smart Education Data System*. Washington, DC: World Bank Group. Ed-DQAF (UIS), UNESCO.

Abdul-Hamid, H. (2014). "What Matters Most for Education Management Information Systems: A Framework Paper." SABER Working Paper Series No. 7. Washington, DC: World Bank.

African Union and ADEA. (2018). Continental Education Strategy for Africa: Indicators Manual. Available at http://www.education2030-africa.org/images/systeme_edu/docs/Final_Indicators_Manual.compressed.pdf.

Community Systems Foundation (www.communitysystemsfoundation.org) and OPENEMIS staff provided comprehensive information on the functionality of OPENEMIS.

Dillman, D., J. Smyth and L.M. Christian. (2014). *Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method*. New Jersey: John Wiley & Sons Inc.

DHIS2. DHIS2 documentation. Retrieved from: https://www.dhis2.org/documentation. Accessed November 2019.

DHIS2. DHIS2 user guide. Retrieved from: https://docs.dhis2.org/2.33/en/user/dhis2_user_manual_en.pdf. Accessed November 2019.

DHIS2. DHIS2 technology platform. Retrieved from: https://www.dhis2.org/technology. Accessed November 2019.

Gartner. (2019). (Analysts) Howson, C., J. Richardson, R. Sallam and A. Kronz. Magic Quadrant for Analytics and Business Intelligence Platforms. Gartner. February 11, 2019.

Gartner. (2019). (Analysts) Ronthal, A., R. Edjlali and R. Greenwald. Magic Quadrant for Data Management Solutions for Analytics. Gartner. January 21, 2019.

Becta. (2006). Functional Requirements for Information Management Including MIS. Retrieved from: http://archive.teachfind.com/becta/industry.becta.org.uk/content_files/industry/resources/Specification_key_docs/IMP_Requirements_final_v1.doc?resID=20860&download_url=/content_files/industry/resources/Specification_key_docs/IMP_Requirements_final_v1.doc. Accessed November 2019.

Heywood, A., and D. Boone. (2015). *Guidelines for Data Management Standards in Routine Health Information Systems*. Measure Evaluation, Carolina Population Center, Chapel Hill, NC.

FHI 360. Factsheet: https://www.fhi360.org/sites/default/files/media/documents/resource-m360-factsheet.pdf.

FHI 360 staff provided comprehensive information on FHI 360 Mobile Architecture.

HISP-South Africa. Health Information Systems Program (HISP) staff provided comprehensive information on the functionality of the DHIS2 system. http://www.hisp.org.

New York State Student Identification System (NYSSIS). (2008). Available at https://www.baruch.cuny.edu/step/documents/nyssisguide.pdf.

OECD. (2007). Glossary of Statistical Terms. OECD: Paris.

OpenEMIS. https://www.openemis.org/.

OpenEMIS. GPE Open EMIS Questionnaire Response. Obtained from OpenEMIS. UNESCO.

OpenEMIS. Lesotho Openemis Pilot. Obtained from Community Systems Foundation (CSF) Technical Team.

OpenEMIS. Overview. Obtained from Community Systems Foundation (CSF) Technical Team.

Stair, R., and G. Reynolds. (2018). *Principles of Information Systems*. Course Technology, Cengage Learning.

Subosa, M., and M. West. (2018). "Re-orienting Education Management Information Systems (EMIS) Towards Inclusive and Equitable Quality Education and Lifelong Learning." Working Paper on Education Policy No. 5, UNESCO.

UNESCO. (2019). Retrieved from: https://learningportal.iiep.unesco.org/en/glossary/educational-management-information-system-emis. Accessed November 2019.

UNESCO Institute for Statistics (UIS). (2019). Retrieved from: http://uis.unesco.org/en/capacity-development-tools. Accessed November 2019.

Van Wyk, C. (2015). "An overview of key datasets in education in South Africa." *South African Journal of Childhood Education,* 5:2. Retrieved from: https://sajce.co.za/index.php/sajce/article/view/394.

Van Wyk, C. (2006). "The Development of an Education Management Information System From a Sensemaking Perspective and the Application of Quantitative Methods to Analyse Education Data Sets," Unpublished PhD Disseration. Stellenbosch University.

Wikipedia. "Software as a Service." Available from: https://en.wikipedia.org/wiki/Software_as_a_service. Accessed December 2019.

Annex A. List of Contributors

This list acknowledges those who contributed by drafting sub-sections of this report; by careful peer review in writing; or by participating in extensive interviews. All are listed here to the best of our knowledge.

Platform	Organization	Name
DHIS2	University of Oslo	Braa, Kristin
DHIS2	University of Oslo	Russpatrick, Scott
DHIS2	University of Oslo	Sanner, Terje Aksel
DHIS2	University of Oslo	Staring, Knut
DHIS2	HISP/South Africa	Claasen, Elmarie
N/A	ADEA EMIS Taskforce	Maiga, Youssouf
N/A	GPE	Irving, Margaret
N/A	GPE	Martinez, Raphaelle
N/A	Microsoft	Joyce, Alexa
N/A	UIS	Boade, Georges
N/A	UIS	Labé, Olivier
N/A	UIS	Montoya, Silvia
N/A	UIS	Sigdel, Shailendra
N/A	Yet Analytics	Blake-Plock, Shelly
OpenEMIS	Community Systems Foundation	Kapp, Jon
OpenEMIS	Community Systems Foundation	Oswalt, Kris
OpenEMIS	OpenEMIS	Chapelete, Pierre
OpenEMIS	OpenEMIS	Turnbull, Karl
OpenEMIS	UNESCO	Bin Mahfooz, Sara
OpenEMIS	UNESCO	Chang, Gwang-Chol
OpenEMIS	UNESCO	Yano, Satoko
SABER-EMIS	World Bank	Azevedo, Joao Pedro
StatEduc	UIS	Fall, Yacine

In addition, the following colleagues participated in peer reviewing this guide verbally and/or by providing advice. Only those who are not already listed above are named here.

Name	Organization
Bah, Alpha	The Gambia - EMIS
Chakroun, Borhene	UNESCO
Dobey, Edward	Ghana – EMIS
Memom, Asif	UNICEF
Myrick, Neal	Tableau Foundation
Ndao, Makha	ADEA
Palacios, Alex	GPE
Reuge, Nicolas	UNICEF
Thompson-Quaye, Anna	GPE

Annex B. List of SDG 4 Indicators and Considerations for EMIS

In this section, the enhanced and changed role of EMIS to collect data for monitoring SDG 4 implementation is emphasized. For EMIS to respond to the increasing demands, it must operate as an integrated and interconnected system with datasets coming from diverse sources. Table B1 highlights how those sources interact with the EMIS "backbone" data, the data that could be in a warehouse that includes the backbone, and/or reports that are typically produced manually, and SDG 4 indicators that are not really "data." The specific relationship between the equity/equality concerns in SDG 4, which are much stronger than they were for the MDGs, is highlighted in the row that pertains to SDG 4.5.1.

Table B.1. SDG 4 Indicators – and How EMIS Could Approach Them

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
Target 4.1: By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes						
4.1.1	Proportion of children and young people (a) in grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex	No.	Yes.	**Interoperability and data integration**: This indicator requires that the EMIS is compatible with the assessment systems. Assessment data integrated, or merged with EMIS data, using a common field (school identifier).	Data on learning outcomes from national or cross-national assessments to use the same school identifier as in EMIS.	Possibly.
4.1.2	Administration of a nationally representative learning assessment (a) in grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education	No.	No.		**Data required unrelated to EMIS**: This is qualitative data on whether such assessments exist and perhaps, in addition, what their parameters or metadata consist of. Data on the administration of a large-scale assessment from a national representative sample from national learning assessment offices, ministries of education or other bodies responsible for learning assessments, including regional or international organizations running learning assessments (e.g. CONFEMEN, EQAP, IEA, OECD, SACMEQ and LLECE).	N/A

1 These could be numerical or qualitative. If numerical, this data would not be in a true database that is linkable to EMIS; they could be in a simple Excel sheet, or even in a Word file, for example.

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.1.3	Gross intake ratio to the last grade (primary education, lower secondary education)	Yes. Typical in an aggregated EMIS (school census approach), but would need to bring in outside population data from census or other sources. Need to keep longitudinally. Quality of data on repeaters may be a limitation. Individual learner tracking could help here.	Yes. May include population data for the denominator.	In an aggregated data collection system, new entrants are calculated by subtracting the number of pupils repeating the last grade from total enrollment in the last grade. In a system where individual learners are captured, the learner's unique identifier makes it possible to follow a learner's progress in the system. Accurate repeaters and dropout totals can be obtained in longitudinal data.		Yes. Difficult to fully automate.
4.1.4	Completion rate (primary education, lower secondary education, upper secondary education)	Yes. For the numerator.	Yes. Population data for the denominator would have to come from census or similar data.			Yes. Difficult to automate.
4.1.5	Out-of-school rate (primary education, lower secondary education, upper secondary education)	Yes. For part of the numerator.	Yes. Population data in denominator typically from census or similar data. Household survey data could triangulate EMIS + census data.	**Interoperability**: Data from EMIS to be linked with population data using a common field.	**EMIS related**: Enrollment by single year of age in each level of education. **Unrelated to EMIS**: Population estimates by single year of age.	Yes. Difficult to automate.
4.1.6	Percentage of children over-age for grade (primary education, lower secondary education)	Yes. Entirely EMIS.	Yes. Could contain household survey data to verify age patterns.			Yes. Difficult to automate.
4.1.7	Number of years of (a) free and (b) compulsory primary and secondary education guaranteed in legal frameworks	No.	No.		This is just a simple policy variable.	N/A

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
Target 4.2: By 2030, ensure that all girls and boys have access to quality early childhood development care, and pre-primary education so that they are ready for primary education						
4.2.1	Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex	No.	Yes. Should have data from household survey or other in the data warehouse.	UNICEF is working on frameworks for measuring this and the methods should start finding their way into household surveys over the period 2021 and onward.		Yes.
4.2.2	Participation rate in organized learning (one year before the official primary entry age), by sex	Yes.	Yes. May need household survey data to amplify the EMIS data. Population data needed for the denominator.	Much may hinge on the definition of "organized" learning that the country uses, which may or may not agree 100% with the conception in the SDG. Could also be amplified via reports from UNICEF-inspired household surveys.		Yes.
4.2.3	Percentage of children under 5 years experiencing positive and stimulating home learning environments	No.	Yes. Would require household survey data.			May need some manual reporting.
4.2.4	Gross early childhood education enrollment ratio in (a) pre-primary education and (b) early childhood educational development	Yes.	Yes. Would require additional population data for the denominator, and would be good to triangulate with household survey data.			May need some manual reporting.
4.2.5	Number of years of (a) free and (b) compulsory pre-primary education guaranteed in legal frameworks	No.	No.		This is just a simple policy variable.	NA
Target 4.3: By 2030, ensure equal access for all women and men to affordable and quality technical, vocational and tertiary education, including university						

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.3.1	Participation rate of youth and adults in formal and nonformal education and training in the previous 12 months, by sex	Yes.	Yes, as EMIS may have a hard time collecting these data. May need to be augmented by data from household surveys, ministry of labor, and other less customary sources of data.			May need some manual reporting.
4.3.2	Gross enrollment ratio for tertiary education by sex	Yes.	In many countries, the data for this may not be in the EMIS backbone at all; tertiary education may be in a different ministry with its own HEMIS. Higher education institutions tend to be defensive about their autonomy. May need to triangulate with household survey data.			May need some manual reporting.
4.3.3	Participation rate in technical-vocational programs (15- to 24-year-olds) by sex	Yes.	Yes, as EMIS may have a hard time collecting these data. May need to be augmented by data from household surveys, ministry of labor, and other less customary sources of data.			May need some manual reporting.
Target 4.4: By 2030, substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship						
4.4.1	Proportion of youth and adults with information and communications technology (ICT) skills, by type of skill	No.	Yes. Would need to be augmented with assessment data. This will be a difficult area to measure. UNESCO is working on methods.	**Interoperability and data integration**: The EMIS is compatible with the assessment systems. Assessment data integrated, or merged with EMIS data, using a common field (school identifier), if relevant (that is, if school-based rather than household or workplace surveys).	Data on learning outcomes from national or cross-national assessments to use the same school identifier as in EMIS. Data may come from household or workplace surveys, though.	Almost certainly.

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.4.2	Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills	No.	Yes. Would need to be augmented with assessment data. This will be a difficult area to measure. UNESCO is working on methods.	**Interoperability and data integration**: The EMIS is compatible with the assessment systems. Assessment data integrated, or merged with EMIS data, using a common field (school identifier), if relevant (that is, if school-based rather than household or workplace surveys).	Data on learning outcomes from national or cross-national assessments to use the same school identifier as in EMIS. Data may come from household or workplace surveys, though.	Almost certainly.
4.4.3	Youth/adult educational attainment rates by age group, economic activity status, levels of education and program orientation	No.	Yes. Would need to be augmented with population (census) or household survey data typically gathered by national statistical offices or ministries of labor.			Yes.
	Target 4.5: 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					
4.5.1	Parity indices (female/male, rural/urban, bottom/top wealth quintile and others such as disability status, indigenous peoples and conflict-affected, as data become available) for all education indicators on this list that can be disaggregated	No. The indices themselves would typically not be produced by EMIS or be part of its backbone.	Yes. It would be good to have these in a warehouse. It would require extraneous data such as census, household survey data, or innovative data sources such as citizen data.	**Interoperability and data integration**: Note that this area makes the SDGs quite different from the MDGs. Aside from that, countries themselves ought to be interested in inequalities and inequities. It is the need to correlate issues such as access, completion, etc., to urban/rural location, wealth, etc., that makes the warehousing and interoperability functions of a modern EMIS so important.		Almost certainly will require much manual input.

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.5.2	Percentage of students in primary education whose first or home language is the language of instruction	No.	Yes. Special household or school surveys.			Yes.
4.5.3	Extent to which explicit formula-based policies reallocate education resources to disadvantaged populations	No.	No.	This is a complex policy variable requiring qualitative information.		Would need to be reported manually.
4.5.4	Education expenditure per student by level of education and source of funding	No.	No.	This a complex policy variable requiring information from the ministry of education or ministry of finance or both.		Would need to be reported manually.
4.5.5	Percentage of total aid to education allocated to least developed countries	No.	No.	This is a complex variable applicable only at the global level.		
Target 4.6: By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy						
4.6.1	Percentage of population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills, by sex	No.	Yes. Would need to be augmented with assessment data and or household or workplace survey data. This will be a difficult area to measure. UNESCO is working on methods.			Almost certainly.
4.6.2	Youth/adult literacy rate	No.	Yes. Would need to be augmented with assessment data and or household or workplace survey data. This will be a difficult area to measure. UNESCO is working on methods.			Almost certainly.

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.6.3	Participation rate of illiterate youth/ adults in literacy programs	Possibly, depending on the degree to which EMIS surveys adult or youth learning centers. These are sometimes nonformal and offered in or by faith-based organizations, community clubs, etc., so it may be difficult.	Yes. Because of difficulties in data collection through traditional administrative means, it may be necessary to use household surveys.			Most likely.
Target 4.7: By 2030, ensure all learners acquire knowledge and skills needed to promote sustainable development						
4.7.1	Extent to which (i) global citizenship education and (ii) education for sustainable development, including gender equality and human rights, are mainstreamed at all levels in: (a) national education policies; (b) curricula; (c) teacher education and (d) student assessment	No.	No, this is a policy variable to be measured through key informants.			Yes, but report is mostly descriptive, less numerical.
4.7.2	Percentage of schools that provide life skills-based HIV and sexuality education	Not traditionally, but could be included in the administrative survey.	Possibly, but would require a schools survey.			Yes. Difficult to automate.
4.7.3	Extent to which the framework on the World Programme on Human Rights Education is implemented nationally (as per the UNGA Resolution 59/113)	No.	No, this is a policy variable to be measured through key informants.			Yes, but report is mostly descriptive, less numerical.

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.7.4	Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability	No.	Yes. Would need to be augmented with assessment data and or household or workplace survey data. This will be a difficult area to measure. UNESCO is working on methods.			Almost certainly.
4.7.5	Percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience	No.	Yes. Would need to be augmented with assessment data and or household or workplace survey data. This will be a difficult area to measure. UNESCO is working on methods.			Almost certainly.
Target 4.a: Build and upgrade education facilities that are child-, disability- and gender-sensitive and provide safe, nonviolent, inclusive and effective learning environments for all						
4.a.1	Proportion of schools with access to: (a) electricity; (b) internet for pedagogical purposes; (c) computers for pedagogical purposes; (d) adapted infrastructure and materials for students with disabilities; (e) basic drinking water; (f) single-sex basic sanitation facilities; and (g) basic handwashing facilities (as per the WASH indicator definitions)	Yes, in some countries. It could be added to the EMIS administrative survey, maybe not annually. One could provide a rolling survey depending on the variables (so that every variable gets covered every three years, for example).	Yes, if there has been a separate schools facilities survey that is not part of the traditional EMIS.			Almost certainly.

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.a.2	Percentage of students experiencing bullying, corporal punishment, harassment, violence, sexual discrimination and abuse	Yes, in some countries. It could be added to the EMIS administrative survey, maybe not annually. One could provide a rolling survey depending on the variables (so that every variable gets covered every three years, for example).	Yes, if there has been a separate schools climate survey that is not part of the traditional EMIS. Or it could be part of a household survey such as MICS or DHS or other.			Almost certainly.
4.a.3	Number of attacks on students, personnel and institutions	Yes, in some countries. It could be added to the EMIS administrative survey, maybe not annually. One could provide a rolling survey depending on the variables (so that every variable gets covered every three years, for example).	Yes, if there has been a separate schools climate survey that is not part of the traditional EMIS. Or it could be part of a household survey such as MICS or DHS or other.			Almost certainly.
Target 4.b: By 2020, substantially expand globally the number of scholarships available to developing countries, in particular least developed countries, small island developing states and African countries, for enrollment in higher education, including vocational training, information and communications technology, technical, engineering and scientific programs in developed countries and other developing countries						
4.b.1	Volume of official development assistance flows for scholarships by sector and type of study	No.	No.	Most likely this would have to be gathered manually from the country's ministry of finance or the budget or donor coordination units of the ministry of education.		Yes.

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.b.2	Number of higher education scholarships awarded by beneficiary country	No.	No.	Most likely this would have to be gathered by a specialized office in the ministry of education or local embassies of other countries.		Yes.
Target 4.c: By 2030, substantially increase the supply of qualified teachers, including through international cooperation for teacher training in developing countries, especially least developed countries and small island developing states						
4.c.1	Proportion of teachers in: (a) pre-primary education; (b) primary education; (c) lower secondary education; and (d) upper secondary education who have received at least the minimum organized teacher training (e.g. pedagogical training) pre-service or in-service required for teaching at the relevant level in a given country, by sex	Yes, in many countries but not all.	In some countries the backbone data could be supplemented in a warehouse with data from the HR or payroll systems and put into a data warehouse.	**Interoperability**: Data from EMIS to be linked with HR or payroll data using a common field.		Almost certainly.
4.c.2	Pupil-trained teacher ratio by education level	Yes, in many countries but not all.	In some countries the backbone data could be supplemented in a warehouse with data from the HR or payroll systems and put into a data warehouse.	**Interoperability**: Data from EMIS to be linked with HR or payroll data using a common field.		Almost certainly.
4.c.3	Proportion of teachers qualified according to national standards by education level and type of institution	Yes, in many countries but not all.	In some countries the backbone data could be supplemented in a warehouse with data from the HR or payroll systems and put into a data warehouse.	**Interoperability**: Data from EMIS to be linked with HR or payroll data using a common field.		Almost certainly.

	Indicator	Central EMIS backbone	Not usually in the EMIS backbone but instead in an EMIS-related data warehouse	Comments	Entirely unrelated to EMIS or even any database[1]	Calculated from EMIS and similar data, but as a manual output
4.c.4	Pupil-qualified teacher ratio by education level	Yes, in many countries but not all.	In some countries the backbone data could be supplemented in a warehouse with data from the HR or payroll systems and put into a data warehouse.	**Interoperability**: Data from EMIS to be linked with HR or payroll data using a common field.		Almost certainly.
4.c.5	Average teacher salary relative to other professions requiring a comparable level of qualification	No.	No.	This is a complex policy variable that is typically derived from household income and expenditure surveys. Salary scales can also help if empirical information is not available.		Yes.
4.c.6	Teacher attrition rate by education level	Not usually, as most EMIS do not track individual teachers, up to now.	Possibly. Could have a warehouse that includes data from payroll or HR and annual differences can be calculated.			Yes.
4.c.7	Percentage of teachers who received in-service training in the last 12 months by type of training	Not typically asked in most EMIS, but it could be.	Possibly. There might be a database of training provided by the relevant wings in the ministry of education or a ministry of higher education. However, much of this training is provided by civil society or the teachers simply procure it for themselves. Thus, adding it as a question in the EMIS backbone would make the most sense.			Yes.

Annex C. Sample EMIS Calendar

An EMIS data collection calendar covers all the steps in the data collection process and should be disseminated to all relevant role players. Below is an example from Ohio, USA.[1]

1 Note that this EMIS calendar is probably much more elaborate than is suitable for lower-income or lower-middle-income countries. Retrieved from: http://education.ohio.gov/getattachment/Topics/Data/EMIS/Reporting-Responsibilities/EMIS-Data-Collection-Calendars/Processing-Schedule-FY19-v0-3-4-29-19.pdf.aspx?lang=en-US or https://tinyurl.com/y9rqmfwt.

EMIS Data Collection Calendar for 2018-19			
By Major Data Grouping	**Data Set**	**Open Date**	**Close Date**
Main Student Collections			
Student Cross Reference (FY19) (2019SSCRS)	S	7/6/2018	7/17/2019
Retention Reporting All Grades (2019SRTNT)	S	7/27/2018	8/31/2018
SOES Beginning of Year Student Collection (FY19) (2019SAODE)	S	8/3/2018	12/20/2018
SOES Student Contact(s) Collection (FY19) (2019SSSDT)	S	8/3/2018	8/9/2019
Beginning of Year Student Collection (FY19) (2019S1TRD)	S	9/4/2018	12/20/2018
Midyear Student Collection (FY19) (2019S2TRD)	S	1/4/2019	4/30/2019
SOES End of Year Student Collection (FY19) (2019SBODE)	S	1/4/2019	7/17/2019
End of Year Student Collection (FY19) (2019S3TRD)	S	5/3/2019	7/17/2019
Additional Student and Staff/Calendar Collections			
Calendar Collection - Initial (FY19) (2019CINIT)	C	7/10/2018	9/28/2018
Calendar Collection - Final (FY19) (2019CFINL)	C	10/2/2018	8/9/2019
Initial Staff and Course Collection (FY19) (2019L1STR)	L	10/2/2018	2/4/2019
March Follow-up Collection (FY19) (2019D0000)	D	2/1/2019	5/10/2019
Final Staff and Course Collection (FY19) (2019L2FNL)	L	2/5/2019	8/9/2019
Graduation FY19 Collection (2019G0000)	G	5/3/2019	10/18/2019

Assessment Collections			
Summer and Fall End of Course Assessment Collection (FY19) (2019AGEFL)	A	9/14/2018	3/29/2019
Summer OGT Assessment Collection FY19 (2019AGXSM)	A	9/14/2018	11/30/2018
Fall Early Learning Assessment Collection FY19 (2019AGBFL)	A	9/28/2018	3/1/2019
Kindergarten Readiness Assessment Collection FY19 (2019AGOFL)	A	9/28/2018	2/1/2019
Child Outcome Summary Assessment Collection FY19 (2019AGMFY)	A	10/15/2018	7/12/2019
Fall 3rd Gr Reading Collection FY19 (2019AGNFL)	A	10/23/2018	3/1/2019
~~Fall~~ DORP Assessment Collection FY19 (2019AGDFL)	A	10/26/2018	7/8/2019
OGT Assessment Collection FY19 (2019AGXFY)	A	1/8/2019	8/9/2019
Other Accountability Assessments FY19 (2019ANACC)	A	2/5/2019	8/14/2019
Spring Alternate Assessment Collection FY19 (2019AALTS)	A	2/5/2019	8/9/2019
~~Spring DORP Assessment Collection FY19 (2019AGDSP)~~	~~A~~	~~1/8/2019~~	~~7/8/2019~~
OELPA Assessment Collection FY19 (2019AGFSP)	A	2/22/2019	8/9/2019
Spring End of Course State Assessment Collection FY19 (2019AGESP)	A	3/1/2019	7/17/2019
Spring Early Learning Assessment Collection FY19 (2019AGBSP)	A	4/9/2019	8/9/2019
Career Tech Accountability Assessment Collection (FY19) (2019ACTSP)	A	3/29/2019	10/11/2019
Financial Collections			
Financial FY19 Collection (2019H0000)	H	6/3/2019	8/30/2019
Financial Supplemental FY19 Collection (2019H0000)	H	9/3/2019	9/27/2019
Five Year Forecast - Initial Required (FY19) (2019P0000)	P	10/1/2018	10/31/2018
Five Year Forecast - Initial Optional (FY19) (2019P1OPT)	P	11/6/2018	3/29/2019
Five Year Forecast - Required Spring Update (FY19) (2019P2MAY)	P	4/2/2019	5/31/2019
Five Year Forecast - Final Optional (FY19) (2019P3OPT)	P	6/4/2019	8/9/2019

Shading indicates date or other changes from prior version of schedule.

Version 0.3
Last Updated 4/29/2019

EMIS Data Collection Calendar for 2018-19			
By Open Date	**Data Set**	**Open Date**	**Close Date**
Student Cross Reference (FY19) (2019SSCRS)	S	7/6/2018	7/17/2019
Calendar Collection - Initial (FY19) (2019CINIT)	C	7/10/2018	9/28/2018
Retention Reporting All Grades (2019SRTNT)	S	7/27/2018	8/31/2018
SOES Beginning of Year Student Collection (FY19) (2019SAODE)	S	8/3/2018	12/20/2018
SOES Student Contact(s) Collection (FY19) (2019SSSDT)	S	8/3/2018	8/9/2019
Beginning of Year Student Collection (FY19) (2019S1TRD)	S	9/4/2018	12/20/2018
Summer and Fall End of Course Assessment Collection (FY19) (2019AGEFL)	A	9/14/2018	3/29/2019
Summer OGT Assessment Collection FY19 (2019AGXSM)	A	9/14/2018	11/30/2018
Fall Early Learning Assessment Collection FY19 (2019AGBFL)	A	9/28/2018	3/1/2019
Kindergarten Readiness Assessment Collection FY19 (2019AGOFL)	A	9/28/2018	2/1/2019
Five Year Forecast - Initial Required (FY19) (2019P0000)	P	10/1/2018	10/31/2018
Calendar Collection - Final (FY19) (2019CFINL)	C	10/2/2018	8/9/2019
Initial Staff and Course Collection (FY19) (2019L1STR)	L	10/2/2018	2/4/2019
Child Outcome Summary Assessment Collection FY19 (2019AGMFY)	A	10/15/2018	7/12/2019
Fall 3rd Gr Reading Collection FY19 (2019AGNFL)	A	10/23/2018	3/1/2019
~~Fall~~ DORP Assessment Collection FY19 (2019AGDFL)	A	10/26/2018	7/8/2019
Five Year Forecast - Initial Optional (FY19) (2019P1OPT)	P	11/6/2018	3/29/2019
Midyear Student Collection (FY19) (2019S2TRD)	S	1/4/2019	4/30/2019
SOES End of Year Student Collection (FY19) (2019SBODE)	S	1/4/2019	7/17/2019
OGT Assessment Collection FY19 (2019AGXFY)	A	1/8/2019	8/9/2019
March Follow-up Collection (FY19) (2019D0000)	D	2/1/2019	5/10/2019
Final Staff and Course Collection (FY19) (2019L2FNL)	L	2/5/2019	8/9/2019
Other Accountability Assessments FY19 (2019ANACC)	A	2/5/2019	8/14/2019

Spring Alternate Assessment Collection FY19 (2019AALTS)	A	2/5/2019	8/9/2019
~~Spring DORP Assessment Collection FY19 (2019AGDSP)~~	~~A~~	~~2/8/2019~~	~~7/8/2019~~
OELPA Assessment Collection FY19 (2019AGFSP)	A	2/22/2019	8/9/2019
Spring End of Course State Assessment Collection FY19 (2019AGESP)	A	3/1/2019	7/17/2019
Spring State Assessment Grades 3-8 Collection FY19 (2019AGNSP)	A	3/1/2019	7/17/2019
Career Tech Accountability Assessment Collection (FY19) (2019ACTSP)	A	3/29/2019	10/11/2019
Five Year Forecast - Required Spring Update (FY19) (2019P2MAY)	P	4/2/2019	5/31/2019
Spring Early Learning Assessment Collection FY19 (2019AGBSP)	A	4/9/2019	8/9/2019
End of Year Student Collection (FY19) (2019S3TRD)	S	5/3/2019	7/17/2019
Graduation FY19 Collection (2019G0000)	G	5/3/2019	10/18/2019
Financial FY19 Collection (2019H0000)	H	6/3/2019	8/30/2019
Five Year Forecast - Final Optional (FY19) (2019P3OPT)	P	6/4/2019	8/9/2019
Financial Supplemental FY19 Collection (2019H0000)	H	9/3/2019	9/27/2019

Shading indicates date or other changes from prior version of schedule.

Version 0.3
Last Updated 4/29/2019

EMIS Data Collection Calendar for 2018-19			
By Close Date	**Data Set**	**Open Date**	**Close Date**
Retention Reporting All Grades (2019SRTNT)	S	7/27/2018	8/31/2018
Calendar Collection - Initial (FY19) (2019CINIT)	C	7/10/2018	9/28/2018
Five Year Forecast - Initial Required (FY19) (2019P0000)	P	10/1/2018	10/31/2018
Summer OGT Assessment Collection FY19 (2019AGXSM)	A	9/14/2018	11/30/2018
SOES Beginning of Year Student Collection (FY19) (2019SAODE)	S	8/3/2018	12/20/2018
Beginning of Year Student Collection (FY19) (2019S1TRD)	S	9/4/2018	12/20/2018
Kindergarten Readiness Assessment Collection FY19 (2019AGOFL)	A	9/28/2018	2/1/2019
Initial Staff and Course Collection (FY19) (2019L1STR)	L	10/2/2018	2/4/2019
Fall Early Learning Assessment Collection FY19 (2019AGBFL)	A	9/28/2018	3/1/2019
Fall 3rd Gr Reading Collection FY19 (2019AGNFL)	A	10/23/2018	3/1/2019
Summer and Fall End of Course Assessment Collection (FY19) (2019AGEFL)	A	9/14/2018	3/29/2019
Five Year Forecast - Initial Optional (FY19) (2019P1OPT)	P	11/6/2018	3/29/2019
Midyear Student Collection (FY19) (2019S2TRD)	S	1/4/2019	4/30/2019
March Follow-up Collection (FY19) (2019D0000)	D	2/1/2019	5/10/2019
Five Year Forecast - Required Spring Update (FY19) (2019P2MAY)	P	4/2/2019	5/31/2019
~~Fall~~ DORP Assessment Collection FY19 (2019AGDFL)	A	10/26/2018	7/8/2019
~~Spring DORP Assessment Collection FY19 (2019AGDSP)~~	A	~~2/8/2019~~	~~7/8/2019~~
Child Outcome Summary Assessment Collection FY19 (2019AGMFY)	A	10/15/2018	7/12/2019
Student Cross Reference (FY19) (2019SSCRS)	S	7/6/2018	7/17/2019
SOES End of Year Student Collection (FY19) (2019SBODE)	S	1/4/2019	7/17/2019
Spring End of Course State Assessment Collection FY19 (2019AGESP)	A	3/1/2019	7/17/2019
Spring State Assessment Grades 3-8 Collection FY19 (2019AGNSP)	A	3/1/2019	7/17/2019
End of Year Student Collection (FY19) (2019S3TRD)	S	5/3/2019	7/17/2019

SOES Student Contact(s) Collection (FY19) (2019SSSDT)	S	8/3/2018	8/9/2019
Calendar Collection - Final (FY19) (2019CFINL)	C	10/2/2018	8/9/2019
OGT Assessment Collection FY19 (2019AGXFY)	A	1/8/2019	8/9/2019
Final Staff and Course Collection (FY19) (2019L2FNL)	L	2/5/2019	8/9/2019
Spring Alternate Assessment Collection FY19 (2019AALTS)	A	2/5/2019	8/9/2019
OELPA Assessment Collection FY19 (2019AGFSP)	A	2/22/2019	8/9/2019
Spring Early Learning Assessment Collection FY19 (2019AGBSP)	A	4/9/2019	8/9/2019
Five Year Forecast - Final Optional (FY19) (2019P3OPT)	P	6/4/2019	8/9/2019
Other Accountability Assessments FY19 (2019ANACC)	A	2/5/2019	8/14/2019
Financial FY19 Collection (2019H0000)	H	6/3/2019	8/30/2019
Financial Supplemental FY19 Collection (2019H0000)	H	9/3/2019	9/27/2019
Career Tech Accountability Assessment Collection (FY19) (2019ACTSP)	A	3/29/2019	10/11/2019
Graduation FY19 Collection (2019G0000)	G	5/3/2019	10/18/2019

Shading indicates date or other changes from prior version of schedule.

Version 0.3
Last Updated 4/29/2019

Education Management Information Systems

Efficiency and Effectiveness in Choosing and Using an EMIS

The Guidelines for Data Management and Functionality in Education Management Information Systems (EMIS) have been produced in partnership between the UNESCO Institute for Statistics (UIS) and the Global Partnership for Education (GPE).

The Guidelines are a direct result of a request from developing countries for guidance on the features that an EMIS should ideally provide, and how to make better use of those features. Country representatives meeting with UNESCO, GPE and other partners felt that since there was no default standard EMIS software produced by the international community, and since it may not be desirable to produce a standard default software, at least a guide of desirable features could be provided. Countries have often produced their own systems or have adopted – partially or in whole – systems provided by various suppliers. In addition, in discussing financial or technical support with development agencies, countries had felt that they do not have a good sense of what they ought to negotiate and discuss with the agencies given the lack of standards or guidelines.

The Guidelines aim to help countries ensure efficiency in choosing and using an EMIS approach. That is, making sure that such systems are as complete as possible while containing as few redundant elements as possible. In addition, the Guidelines emphasize that the systems ought to make it easier to provide value-added information to policy makers and managers by maximizing the interoperability and linkages between the various databases that education sectors manage. For instance, allowing for the identification of schools that might be underperforming relative to their circumstances, or performing better than expected and thus providing lessons learned. Finally, partners requested that the Guidelines provide a holistic view of what a well-functioning EMIS system should look like, in terms of connections to an education sector's stated goals as well as flows and links between data elements (enrollment, cost, personnel, etc.) in the sector. The Guidelines are framed within the context of the fourth Sustainable Development Goal for Education (SDG 4) without implying that global reporting ought to be the main aim of EMIS systems. Some countries have also requested information and guidelines on how one could implement learner and teacher identification systems so as to improve the tracking of learners who drop out and hence school completion. These guidelines strive to fulfill all the aforementioned missions. The UIS and GPE hope that countries and partners find the Guidelines useful in fulfilling their requests.

United Nations Educational, Scientific and Cultural Organization
UNESCO INSTITUTE FOR STATISTICS
SUSTAINABLE DEVELOPMENT GOALS
GPE Transforming Education