EdTech Hub

Clear evidence, better decisions, more learning.

Unlocking Data

GUIDANCE NOTE

Education Data Mapping in Sub-Saharan Africa

Moving from theory to practice

Date May 2022

Authors Irene Selwaness
Taskeen Adam
Laté Lawson
Lucy Heady

DOI 10.53832/edtechhub.0096

UKaid from the British people

THE WORLD BANK

unicef for every child

About this document

Recommended citation

Selwaness, I., Adam, T., Lawson, L., & Heady, L. (2022). *Guidance Note on Education Data Mapping in Sub-Saharan Africa: Moving from theory to practice*. Technical Report . https://doi.org/10.53832/edtechhub.0096. Available at https://docs.edtechhub.org/lib/KQUDWCXC. Available under Creative Commons Attribution 4.0 International, https://creativecommons.org/licenses/by/4.0/.

Notes

EdTech Hub is supported by UK aid (Foreign, Commonwealth and Development Office), Bill & Melinda Gates Foundation, UNICEF and the World Bank. The views expressed in this document do not necessarily reflect the views of UK aid (Foreign, Commonwealth and Development Office), Bill & Melinda Gates Foundation, UNICEF and the World Bank.

This note was produced in partnership with Laté Lawson and Lucy Heady from Education Sub Saharan Africa (ESSA), a charity using evidence to improve education in sub-Saharan Africa.

Reviewers

Molly Jamieson Eberhardt, Eldah Onsomu

Acknowledgements

We are grateful for the case study insights contributed by Esme Kadzamira, Iman Beoku-Betts, Chris McBurnie, and Eldah Onsomu, without which this guidance note would not have been possible.

Contents

Figures and tables

Abbreviations and acronyms

CERT	Centre for Education Research and Training
CGD	Centre for Global development
EMIS	Education Management Information System
ERF	Economic Research Forum
HEI	Higher Education Institution
KIHBS	Kenya Integrated Household Budget Survey
KIIs	Key Informant Interviews
MANEB	Malawi National Examination Board
MODESA	Malawi Open Data for Education Systems Analysis
MoU	Memorandum of Understanding
NEET	Not in Education, Employment or Training
NGO	Non-governmental organisation
SDG	Sustainable Development Goal
TVET	Technical and Vocational Education and Training

1. Introduction

Over the past two years, Unlocking Data[1] has endeavoured to support its community of practice to 'unlock' education data to increase its availability and use to improve education. Through a series of workshops run by Unlocking Data, the need to map the status of access and usage of education data in Africa was identified in order to determine where the major data gaps were. Around the same time, a series of initial data mapping initiatives sprung up in a handful of African countries, all with different purposes, stakeholders and methodological approaches. This guidance note draws lessons from education data mapping experiences in Kenya, Malawi, and Sierra Leone, building on the conceptual framework of education data mapping methodology by Lawson & Heady (2021), and key informant interviews (KIIs).

In this note, we navigate the different questions and purposes, challenges and key considerations that were learnt from these mapping initiatives in order to make them successful and beneficial for other countries and stakeholders intending to pursue similar exercises.

Such mapping exercises can be extremely helpful to education decision-makers and researchers for various reasons. They contribute by:

1. Identifying and listing what each institution collects and produces in terms of education data;
2. Identifying what type of education data each institution needs for its policymaking;
3. Making rich administrative and microeconomic datasets that are nationally collected available in a secure way;
4. Bridging the gap in the utilisation of rich datasets;
5. Conducting in-depth analysis that can inform policy in specific educational domains and help prevent inappropriate policy decisions based on inaccurate research.

The key questions to think about when starting a data mapping project are:

1. What are the different purposes driving data mapping? E.g.,
    - Is it about understanding the data ecosystem?

[1] See https://edtechhub.org/2021/11/11/the-story-of-unlocking-data/

   - Is it about identifying datasets that can be available and accessible for use and / or making those datasets readily available on a portal for researchers?

2. Who are the potential stakeholders in the process and who are the end-users for the data mapping exercise?
   - Is it the government, the research community, and / or others?
3. What are the intended products?
   - Is it a report / research paper, an archive of datasets, a platform / dashboard, and / or other products?
4. What is the area of focus in education and the types of data covered?

The key challenges we identify as related to effective data mapping are as follows:

1. **Data issues** — Accessibility is limited. There is also a lack of technical skills for data analysis which leads to a lack of demand for secondary data. There are also trust issues in making data accessible.
2. **Stakeholders** — There are many stakeholders in some cases. Bringing them together is not straightforward.
3. **Presentation and dissemination of the mapping exercise output** — How to combine and present the output that has been produced in a user-friendly format: is it a question of using a presentation or report, creating a platform or updating an existing one? The data cataloguing platform is new territory.
4. **Resources** — The process may take a long time, so hiring and maintaining the right team is challenging.

Our guidance discusses some of the key concepts and considerations — the three Cs — that are important to the data mapping exercise and can contribute to easing those challenges.

1. Establish a **communication** strategy:
   - Define the end-user(s) and engage with them early;
   - Collaborate (especially with the government); develop trust and buy-in;
   - Communicate on the importance of secondary data in robust evidence-based policymaking;

   - Establish clear communication on the 'how' of the platform / data catalogue.

2. Leverage **collective** wisdom:
   - Have patience
   - Understand it is a process
   - Tread carefully.
3. Develop, learn from, and engage with the **community** of practice.

This note is structured around each of these areas as follows:

- Following this introduction, Section 2 provides a summary of the three data mapping projects that this guidance note is based on.
- Section 3 discusses in detail the purposes behind data mapping initiatives and the questions to think about when starting it.
- Section 4 examines the main challenges that these initiatives faced.
- Section 5 moves on to the key concepts and steps that are important to consider in data mapping processes.
- Section 6 discusses the usefulness of this exercise in other geographies.
- Finally, Section 7 presents the way forward and concludes the report.

2. Case studies

This guidance note is based on the following three case studies:

1. Mapping of TVET data in Kenya
2. Mapping of administrative data in Malawi
3. Mapping of the supply and demand for education data in Sierra Leone

2.1. Mapping of TVET data in Kenya

This project started as part of the Unlocking Data initiative.[2] To ensure quality education as part of Sustainable Development Goal 4 (SDG-4), there is an immediate need to assess numerous education-related targets and indicators that are not usually available or accessible.[3] Thus, this project aims at bridging the data and evidence gap in Technical and Vocational Education and Training (TVET) in Kenya, to contribute to achieving SDG-4 (↑Ministry of Education, 2021).

The project aims at identifying datasets related to TVET, along with their relevance and quality, as well as the actors involved in the data collection / production and use. Five thematic areas were adopted to guide the exploration of the data ecosystem and value chain. These include access and equity, quality, relevance, resources and utilisations, and governance and management issues. The project adopted two methods in the process of the identification of education data, based on ↑Lawson & Heady (2021):

1. The first method is through an online search for all education datasets by public institutes of statistics, national and international organisations, as well as data collected by Higher Education Institutions (HEIs), non-governmental organisations (NGOs), private providers of education, and researchers.
2. The second method is screening empirical research for education data.

The data gaps were identified across the five thematic areas by narrowing down the specific indicators and the data sets required for each thematic area. The process involved identifying the existing datasets and sources by linking them to demand and supply, and also establishing the data gaps across the

[2] For further reading on this initiative, see ↑Adam et al., 2020; ↑Lawson & Heady, 2021; ↑McBurnie & Beoku-Betts, 2021; ↑McBurnie et al., 2021; ↑Pambe et al., 2021; ↑Unlocking Data, 2021.
[3] https://sdgs.un.org/goals

thematic areas. Further, the project documented the mapped documents, studies, and data in line with the thematic areas (↑Ministry of Education, 2021).

Following the identification of datasets, the project will hold country-level consultations with stakeholders to discuss the data gaps and what data is needed for a clearer picture of the TVET system (↑Ministry of Education, 2021).

2.2. Mapping of administrative data in Malawi

This project started as part of Malawi Open Data for Education Systems Analysis (MODESA) led by the Centre for Education Research and Training (CERT) at the University of Malawi in partnership with the Centre for Global Development (CGD). The main aim of the project is to retrieve and use administrative data and microeconomic datasets (i.e., secondary data[4]) for advanced analysis beyond descriptive work that exists in the bulletins and reports released to the public. In-depth analysis of these datasets will bridge the gap in evidence around education and will help to better inform decision-making by the relevant ministries in the education sector (↑Pambe et al., 2021).

The datasets that the project covers are:

- Education Management Information System (EMIS) data at the school level;
- Malawi National Examination Board data at the students level;
- Cohort tracking data on primary students that the government has been collecting since 2016;
- Several (4–5) waves of household survey data that has some information on education;
- Data on public universities from the Ministry of Higher Education.

Once these datasets are made available, they will be used in producing two research papers on education.

In a further phase, this project aims to create an open data portal that acts like a data repository where datasets and relevant documentation (codebook, questionnaires, etc.) can be publicly accessible (with agreed terms and

[4] Primary data is data that is collected by a researcher or team of researchers, for a specific research purpose or analysis under consideration. Secondary data is collected by someone else, including but not limited to national statistical offices, research institutions, etc. There are many examples of secondary data such as, but not limited to, administrative data, household surveys, labour market surveys, population censuses, etc. (↑Hox & Boeije, 2005; ↑UNESCO, 2021; ↑Vartanian, 2011).

conditions of use). The objective of this platform is to encourage researchers, scholars, development partners, civil society, and students to use secondary data. This is because in Malawi, there is reluctance towards using secondary data, as handling the datasets in question and making them available for use (cleaning, harmonising the variables, and potentially compiling them with necessary variables from other datasets to complement the analysis) is quite a difficult process. This is why there is a higher preference for collecting primary data than using secondary data collected by the national statistics offices and relevant ministries (↑Pambe et al., 2021).

2.3. Mapping of the supply and demand for education data in Sierra Leone

This project began as part of engagement by the Ministry of Basic and Senior Secondary Education (MBSSE) in the use of data for decision-making. It aimed at mapping the supply of and demand for education data. This project was initiated as a result of the work between EdTech Hub and the Government of Sierra Leone. A lot of the work between the Hub and the government was somehow being blocked or the potential impacts have been limited because of significant levels of duplication and fragmentation across the system. This duplication or fragmentation means that different organisations may have been collecting the same datasets or handling data on the same topics, but they were not necessarily talking to each other. As a result, there was no 'single source of truth'. Such lack of coordination between data sources and fragmentation of data undermined the government initiative on data being critical to policy development / reform. This project, therefore, was born out of a need to understand the data ecosystem by identifying who collects, handles, uses, and needs what data to support education decisions in Sierra Leone. Such mapping was based on an iterative process of consultation with education stakeholders, involving key informant interviews with government officials, district-level representatives, and donors / other actors (↑McBurnie & Beoku-Betts, 2021).

This mapping process will identify steps to further consolidate the ecosystem, link different data sources, reduce fragmentation, and move towards a 'single source of truth' to inform the design, implementation, and impact of education policies and programmes.

3. Key questions

As discussed in the introduction, when starting data mapping, it is important to think about and try to answer some key questions, such as:

1. What are the different purposes driving data mapping?
2. Who are the potential stakeholders in the process and who are the end-users for this data mapping exercise?
3. What is the area of focus in education?
4. What are the types of data that the data mapping intends to cover?

Key Question 1: What are the different purposes driving data mapping?

The objectives behind data mapping vary from one country to another. They can include the following purposes but they are not mutually exclusive:

a. Understanding the data ecosystem, which entails identifying who the stakeholders are, what data they collect, and the types of indicators they have.

b. Identifying the needs of the different national institutions in terms of data to better inform their education policies, design, and implementation.

c. Identifying, accessing, or generating a list of datasets that can be available and accessible for (re)use.

d. Generating a list of data indicators that can also be available and accessible for use.

e. Identifying the most recent studies on education and the accessibility of data these studies used.

f. Making datasets readily available on a portal — for researchers, among others.

Table 1 maps the different purposes of data mapping to our three countries of focus and discusses potential users for each type of purpose.

Table 1. *Types of objectives by country. Source: Authors' own, based on KIIs.*

Objective / Scope	Countries in the case studies mentioned in this guidance note	Potential end-user
a. Understanding the data ecosystem	Kenya, Sierra Leone	The government and other institutions involved in education policymaking
b. Identifying the needs of the different national institutions involved	Sierra Leone	
c. Identifying, accessing, or generating a list of datasets	Kenya — (identification and listing what is publicly available and accessible) Malawi — (gaining access to some datasets)	In addition to the government, data mapping can also serve the research community and practitioners who may use those datasets or data indicators to conduct data analysis and studies
d. Generating a list of data indicators	Kenya	
e. Identifying the most recent studies on education and the accessibility of the data used	Kenya	
f. Making datasets readily available on a portal — for researchers, scholars, development partners, civil society, and students	Malawi (second phase of the project)	

Key Question 2: Who are the potential stakeholders in the process and who are the end-users for this data mapping exercise?

Who the end-users of this exercise are depends on the objectives of the mapping, as illustrated in Table 1 above.

If the project aims to address objectives 'a' or 'b', then the end-users are usually the government and those institutions involved in education policymaking.

If the objectives extend to cover objectives 'c' to 'f', then it can also be targeted to the research community and practitioners who may use those datasets or data indicators to conduct data analysis and studies.

Key Question 3: What is the area of focus in education?

It is important to determine the education level or area of focus while conducting the data mapping. For instance, TVET was the area of focus for the data mapping exercise in Kenya. Primary and secondary education was the area of focus for the data mapping exercises in Malawi and Sierra Leone.

Key Question 4: What are the types of data that the data mapping intends to cover?

Figure 1 below presents different types of data that might be collected depending on the objectives of the data mapping process.
The data mapping approach for Objectives a and b was to:

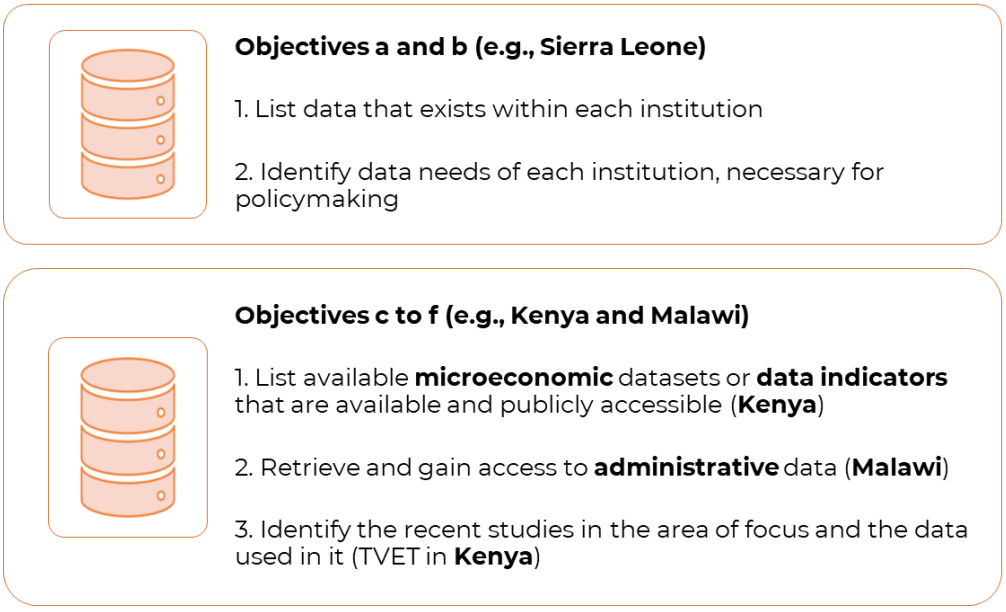
- Examine what each institution collects and produces in terms of data (e.g., Sierra Leone). This approach was about listing the data that exists within each institution, without examining its accessibility / availability, nor aiming to retrieve the datasets / data indicators.
- Identify what type of data each institution needs for its policymaking.
- As for Objectives c to f, the data mapping approach may aim to:
    1. List the available **microeconomic** datasets, e.g.,
        - Kenya Integrated Household Budget Survey (KIHBS)
        - Cross-sectional survey on those not in education, employment, or training (NEET)
        - Surveys of youth skills in formal / informal sectors

2. List the available data indicators that can be publicly accessible, while remaining accurate, such as in Kenya.
3. Gain access to **administrative** data (e.g., Malawi National Examination Board) or **microeconomic** data (e.g., multiple waves of household survey data that has some information on education) for research purposes that can guide and inform education policymaking, such as in Malawi.
4. Negotiate to access databases that are not open or publicly available through the removal of modules that can contain sensitive information (e.g., Malawi).

Some of the data mapping approaches, such as in Kenya on TVET data mapping, also involved identifying the recent studies in the area of focus and the data used in it. In this exercise, the data mapping approach is limited to studies that used available and accessible data, because of issues of accuracy and representativeness. The goal is that such data can be reused by the government, practitioners, or research community, so there must be a condition that such data is accurate and leads to robust evidence, rather than raising issues of credibility. Other datasets collected by researchers or students — i.e., primary data — that are not representative, are excluded.

It is important to note that it is possible that data types covered in the data mapping process evolve along the mapping process in order to meet the expectations of the government, while still meeting researchers' expectations. This is because government needs, in terms of data, are sometimes different from researchers' needs..

Figure 1. *Types of data mapped. Source: Authors' own, based on KIIs.*

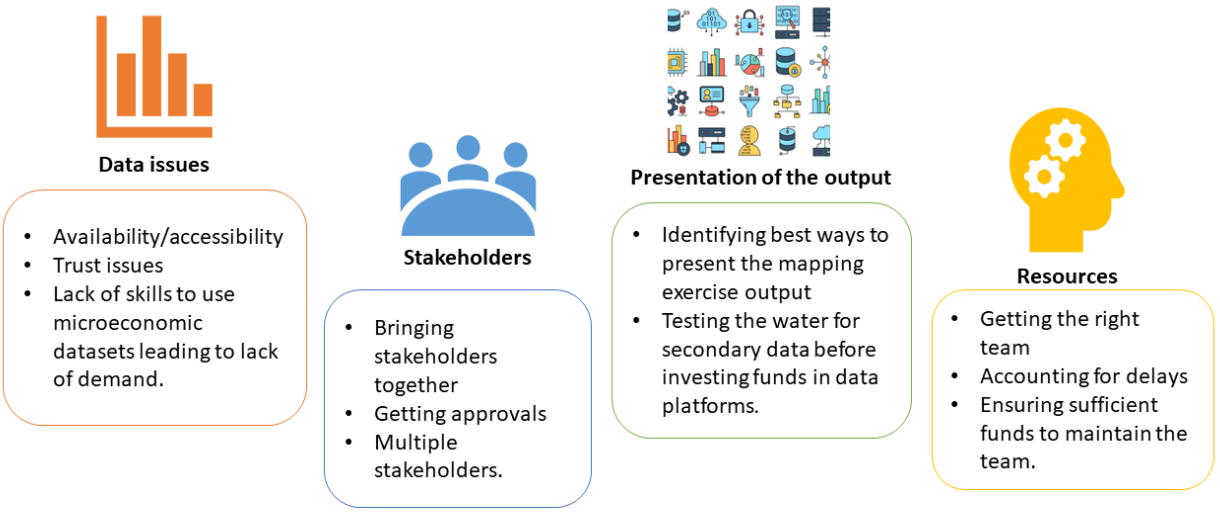


4. Key challenges

The main challenges in data mapping projects can be summarised around the following interlinked topics:

1. Data issues
2. Stakeholders
3. Presentation and dissemination of the output
4. Resources

Figure 2. *Summary of key challenges for data mapping projects. Source: Authors' own, based on Section 4.*



4.1. Data issues

Data accessibility is limited. There is also a lack of technical skills for data analysis, which leads to lack of demand for the use of secondary data. Besides, there are trust issues in making data accessible.

4.1.1. Data availability and accessibility

There is variability in countries' experiences around data availability and accessibility. In some contexts, the government is not likely to share administrative or even census data. In other contexts, the government might state that they can share such data provided sensitive personal information is

removed from databases. However, access to data is still unsuccessful and problematic.

4.1.2. Lack of trust

One of the root causes behind issues of data accessibility is lack of trust and other political reasons. Data owners may be uncomfortable about what the data may be used for. There is a real fear of data users unearthing information that data holders do not want to expose or reveal in the public domain.

4.1.3. Lack of skills in using microeconomic datasets leading to lack of use

Once available data is mapped, effective use of the data is the next issue. This is mainly because potential users may lack the required technical skills for data analysis and academic writing.

In order to increase the use of microeconomic data sets, capacity needs to be built for conducting data analysis, starting from the construction of relevant indicators (even the basic ones) to the effective presentation of findings to policymakers (in policy briefs, presentations, workshops, meetings, etc.).

Figure 3. *Implications of lack of skills in the use of microeconomic datasets.*

The lack of adequate technical skills for using microeconomic data sets has two implications. First, such gaps in technical skills alongside data availability and accessibility issues mean there is limited empirical research on specific areas of education. For example, in Kenya, empirical research on TVET using secondary data (either administrative data such as data from EMISs, or school census or microeconomic datasets such as the KIHBS) has been scarce. Further, most articles that the team screened and found came from the Kenya TVET journal and very little is published in international academic journals.

Second, the lack of skills has contributed to the low utilisation of data in decision-making and has reinforced the weak links between policymaking and data evidence for education. The limited ability to use and analyse data has been identified as the major constraint behind reform of EMISs, as seen in many countries that improved the collection processes of their EMISs but stumbled when it came to using the data (↑Rossiter, 2020).

4.2. Stakeholders

In some cases, many stakeholders are involved. Bringing them together is not straightforward.

4.2.1. Bringing stakeholders together

One of the main challenges that was repeatedly highlighted was about bringing together stakeholders (data producers and holders, including public actors involving relevant ministries or statistical offices and non-state actors) or arranging to meet the different (and multiple) stakeholders. This was true to various degrees in the three countries of focus, but to a lesser extent in Malawi (See Figure 4 for more details on the Malawi experience in this regard).

4.2.2. Getting approvals

Getting the necessary permissions and approval from the government to start the data mapping process is often a necessary prerequisite that can stagnate progress. Additionally, difficulties convening meetings can lead to further delays.

4.2.3. Multiple stakeholders

The area of focus / education level, as for example in TVET, is very dynamic and constantly changing. The ecosystem is not compact, but rather widespread, and consists of many institutions, whether public or made up of non-state actors such as the private sector, or NGOs, as well as different levels and players. For this reason, the process can be more challenging as data is scattered across counties and institutions. What is more, as is the case in Kenya, reaching out to private education institutions (as well as retrieving their data) is a problematic part of the data mapping process. This is related to issues of fragmentation and questions on how representative the data of the TVET system is. This is why starting with mapping data from public institutions seemed a sensible choice in Kenya.

Figure 4. *Foundation of partnerships / relationships with the government.*

There is a historical link between the CERT at the University of Malawi and the Ministry of Education. The centre was established in the early 1990s as an initiative of the Ministry of Education and the University of Malawi to provide the ministry with evidence-based research findings that could help formulate public policies. The director of planning within the Ministry of Education is a member of the board that runs the CERT. This made government buy-in relatively easier.

4.3. Presentation and dissemination of the output

How to combine and present the output that has been produced in a user-friendly format is an important question. Is it one of using a presentation or report, creating a platform, or updating an existing one? The data cataloguing platform is new territory.

4.3.1. Identifying the best ways to present the output of the mapping exercise

At the current phases of the three projects of focus, most of the dissemination or presentation of output is being undertaken through reports. For projects such as that in Sierra Leone, aiming to understand the data ecosystem —who produces what and who needs / uses what in terms of data in their decision-making — questions have arisen about the best way to present that ecosystem. According to a representative from the Sierra Leone team:

> *"We had quite a lot of discussions about just working out how to present this information in a way that people are really able to understand it and absorb it. And we did spend quite a lot of time doing that as well, which was obviously beneficial."* (Sierra Leone team representatives, 2022)

4.3.2. Platform / data cataloguing and repository

For the projects that aim to identify and gain access to data, the future phases for presentation of the output may include:

- Establishing a cataloguing system or building on an existing one. This cataloguing system will serve as a portal for researchers to browse, search, compare, apply for access, and download relevant survey microdata (e.g., Malawi).
- Establishing an archive repository for data indicators and / or empirical research papers following a metadata standard to describe those data indicators or those empirical papers (e.g., Kenya).

4.3.3. Testing the water for secondary data before investing funds in data platforms

Prior to building or updating an existing platform or data catalogue that serves as a portal for researchers to browse, search, compare, and apply for access, both the research team engaged in the data mapping and the donors first need to demonstrate:

1. How the secondary data will be used in practice;
2. Whether the data can really be made accessible before investing their funds in establishing a data platform or updating an existing one.

> *"There are plans for the data platform, but we didn't go into depth … and we stopped thinking about the database. We are more concentrating on how we're going to be able to retrieve that data that we want."* (Malawi team representative)

Although no projects have reached this phase yet, creating a data cataloguing platform or updating an existing one and opening up the data to the public is considered one of the most intensive future parts of the process. It is new territory and a learning process that needs support on many fronts: knowledge of legal issues regarding data sharing, staffing the right persons (IT specialists, system analysts, data managers, statisticians, etc.), ensuring the right skill sets, and maintaining adequate funding.

4.4. Resources

The process may take a long time, so hiring and maintaining the right team is challenging

4.4.1. Getting the right team

It is crucial to have a team composed of researchers and statisticians who can understand and investigate what data is there.

To speed-up approvals and negotiations, it is also important to have members of the team working with or in connection with the government.

4.4.2. Accounting for factors that could lead to delays

The process of getting approval letters from the government causes delays.

The process of arranging meetings with different public officials / departments is also challenging and time-consuming.

Covid-19 has also delayed some of the teams in maintaining timely negotiations with the government or public institution in charge.

4.4.3. Ensuring sufficient funds

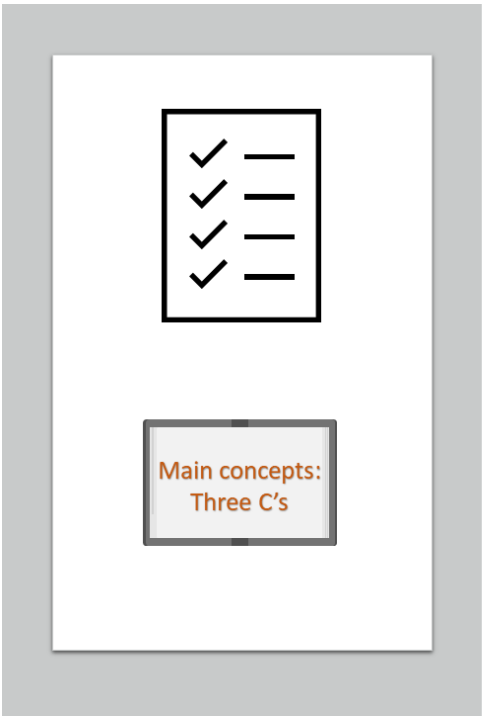
Due to the above-mentioned types of delays, challenges often arose in project planning / management over the life cycle of a project. Thus, the main challenge was how to ensure sufficient funds to maintain the team who have the technical and institutional knowledge of the project involved (e.g., Kenya).

5. Key concepts

Establishing an **enabling environment** for data mapping is a key factor for success in data mapping. There are some fundamental concepts to follow, learn, and develop to establish and sustain this enabling environment for data mapping. These concepts can be summarised around the three Cs:[5]

1. Communication strategy / approach
2. Collective wisdom
3. Community of development partners / practitioners

Figure 5. *Important concepts to consider in data mapping efforts. Source: Authors' own.*



Communication approach

- Identify who you are going to engage with first
- Collaborate
- Communicate the importance of secondary data
- Communicate on the 'how' of the platform/data cataloguing

Leverage **collective** wisdom

- Be patient
- Understand it is a process
- Tread carefully

Develop/learn from the **community** of practice

- Engage with development partners

5.1. Communication strategy / approach

The data mapping experiences in Kenya, Malawi, and Sierra Leone mostly focused on national datasets and thus the main stakeholders were public institutions, such as relevant ministries, directorates, or national statistics offices. The points below, while mostly giving examples of how to establish and maintain good communication with the government, can also apply to non-state actors.

[5] This concept is borrowed from Marie Assaad's fundamental concepts of social capital, the 3 Cs (↑Spencer, 2016), and adapted to the context of key considerations for data mapping.

5.1.1. Identify who you are going to engage with first

It is crucial to identify the right person, institution, or directorate within the government / ministry / national statistics office with whom the team should first engage and present the project to. Experiences that started at the very top of the ministry were likely to get support from there, before moving down to the remaining levels / directorates. A prerequisite for this step is to start small with a clear objective.

5.1.2. Collaborate

In all countries, it is important to establish a foundation of relationships and engage in collaboration with the institution in charge — whether relevant ministries, national statistics offices, or other actors — for purposes of data access, capacity building, and understanding data needs.

In Malawi, a key factor that encouraged the government to share the Malawi National Exam Board (MANEB) data was the involvement of a researcher from within the national examination board and the data manager of MANEB in the data mapping exercise. This was a key factor in reaching the memorandum of agreement between CERT and MANEB to retrieve the data.

Collaboration serves as capacity building for researchers and offices from within the national institutions who produce data so that they learn how to use and analyse those datasets in depth. It would also enable them to make use of some of the underutilised data (due to a lack of technical skills). An example of such data is the cohort tracking data in Malawi that the ministry has been collecting annually since 2016 but that has not been analysed. This capacity-building component can also apply to non-state institutions that are involved in collecting data.

Moreover, through collaboration, the team from the government will be able to review the study results together with the institution in charge of data mapping, while building trust amongst all parties.

Another important consideration is to try in all data mapping exercises to understand what governments need in terms of data / analysis and put that forward.

5.1.3. Communicate and demonstrate the importance of opening up secondary data for the purposes of robust evidence

Secondary data, or data that is being collected by national statistics offices, ministries, or non-state institutions, often presents robust evidence for policymaking (↑Rossiter, 2020; ↑Siddiqui, 2019). This is not the case for primary data, which can often lack representativeness or is designed for a specific

research purpose, and raises questions of generalisation and validity. It is usually difficult to convince policymakers and relevant stakeholders with results that are based on primary data. They often argue that these results cannot be generalised, are not representative, and capture only a small proportion of the reality. Therefore, making secondary data accessible for in-depth data analysis is crucial for robust evidence that is necessary to inform policymaking.

5.1.4. Clearly communicate the 'how' of the platform / data cataloguing

For this to happen, it is necessary to give the government details on how its data would be accessible to the public. This includes discussing all sorts of user agreements, confidentiality agreements, anonymisation processes, citation protocols, etc. This is especially crucial in countries with no data sharing policies. It is important to discuss with data producers if entire datasets would be made publicly available or only specific parts of them. There can also be agreements to make the data less sensitive by removing any confidential information, or any modules / parts that can lead to a potential breach of protocols / confidentiality. This is coupled with providing examples of other data that have been shared e.g., the Data First[6] initiative in South Africa and other African countries, the Economic Research Forum's (ERF) data portal [7] in the Middle East region,[8]and IPUMS[9] that contains the world's largest collection of census microdata covering over 100 countries, and so on.

5.2. Leverage collective wisdom

5.2.1. Be patient

Meeting stakeholders and retrieving data is a long process. It is also crucial to have patience in arranging meetings with stakeholders.

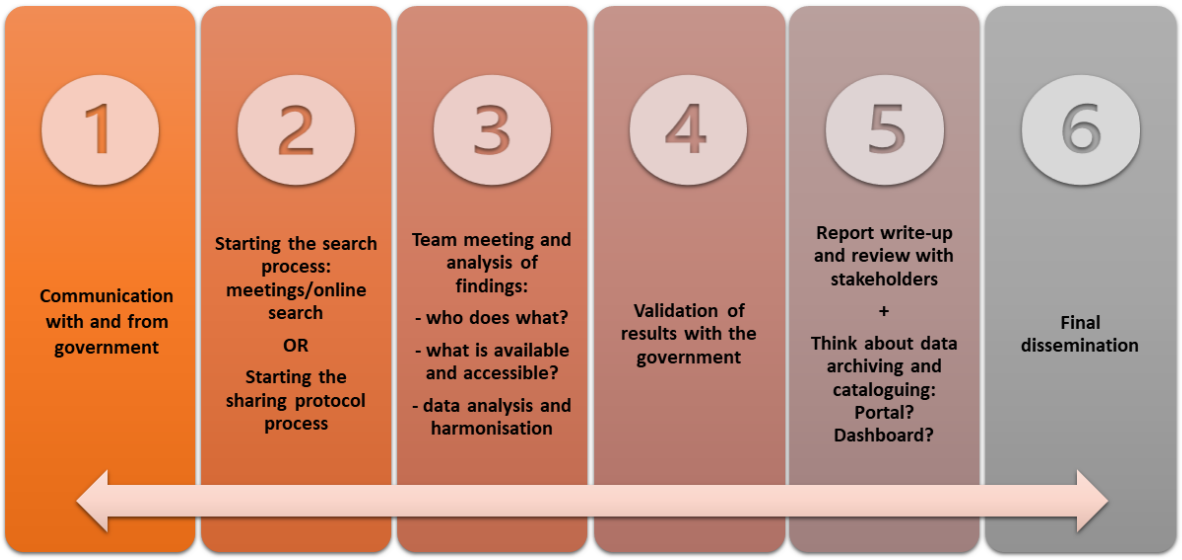
> *"At first, I expected that it would be so easy to retrieve the data, but I found it's a long process even with government data."* (Kenya team representative, 2022)

[6] See https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/central
[7] See http://www.erfdataportal.com/index.php/catalog
[8] These data catalogues are based on the NADA microdata cataloguing tool. This is an open-source microdata cataloguing system, compliant with the Data Documentation Initiative (DDI) and Dublin Core's RDF metadata standards. It serves as a portal for researchers to browse, search, compare, apply for access, and download relevant census or survey datasets, questionnaires, reports, and other information. See https://nada.ihsn.org/ for more details.
[9] See https://international.ipums.org/international/

5.2.2. Understand it is a process

It is also important to understand that data mapping is a new area for many governments in Africa, so there is a need to build trust and walk governments through the process. Figure 6 illustrates the different steps in the data-mapping process around the three selected countries: Kenya, Malawi, and Sierra Leone. This is based on the current phases of the data-mapping process. It does not reflect or include any next or future steps.

Figure 6. *Illustrative data-mapping process in the current phases of the case studies. Source: Authors' own.*



5.2.3. Tread carefully

Taking the above-mentioned challenges into consideration, namely the lack of trust that can lead to issues around data access, it is crucial for data mappers to build trust with data holders so that research can work for the good of the public and hand-in-hand with public policies.

To achieve this, data mappers / analysers have to ensure that data producers / holders are familiar with their research results and with the tone used to disseminate findings.

For example, a collaborative mode of working with the government, as in the case of Malawi, discussed above, ensures that research findings (based on government-sourced data) are presented to and discussed with the government in question before they are made public. This guarantees that results do not create unwanted or unexpected tensions, which can often

happen if the press is involved in disseminating findings and escalating related issues.

> *"When you want to do data mapping, collaboration is the key, particularly if you are using secondary data, i.e. other data producers' data, not your own."* (Malawi team representative, 2022)

5.3. Develop, learn from, and engage with the community of practice

It is crucial to engage with and learn from development partners with similar experiences in data mapping. These can include but are not limited to, experiences in establishing data access protocols, setting-up data cataloguing systems, developing an EMIS, or understanding the data ecosystem. Future steps might include drawing on similar initiatives, e.g., Data First, ERF data portal, and IPUMS.

6. Usefulness of repeating the data mapping exercise

> **"This exercise is required yesterday … It is the engine of everything."**
>
> **- Kenya team representative, 2022**

6.1. Bridging the gap in utilisation and awareness

It is becoming clear that most of the data on education that is collected nationally, in the three countries of focus and perhaps in other SSA countries, has seldom been analysed and has only been used for descriptive reports. Yet, this data is very rich and can be used in informing policy in specific educational domains. Also, many data producers are not aware of the extent of in-depth analysis that can be undertaken using such rich datasets. In-depth analysis of education data can also prevent inappropriate policy decisions based on inaccurate research (↕Woolfrey, 2013), and enhance *information governance*.

Interestingly, relevant ministries or data producers were impressed when they learnt about other countries' experiences in analysing similar datasets and their usefulness for evidence-based policymaking. It stimulated positive reactions as to the breadth of analyses that can be done using similar data. Hence, their attitudes vis-à-vis collaborating with research centres to share and analyse the data may change. An inspiring example from Malawi is presented below:

> *"Most of the data at the exam board is never analysed beyond just reporting the results, and yet it's a rich dataset that can be used to actually inform policy, in particular in relation to the performance of students. (…) When we've engaged with MANEB staff later on, and showed them what Kenya has done, they were actually impressed to see that they do this — the Kenyan examination board — as we got hold of several articles based on analysis that Kenyan examination board has done, with stakeholders within Kenya. They were actually impressed, the officers there (at MANEB) were not aware that this can be done with examination data."* (Malawi team representative, 2022)

6.1.1. Better advocacy

Evidence from Sierra Leone indicates that the mapping of the data ecosystem has contributed to better advocacy for what is needed and to reach agreements on areas of development, such as having a consolidated source of truth.

7. Way forward and concluding remarks

This section discusses the planned next steps for each of the case studies and concludes with main takeaways on education data mapping in sub-Saharan Africa.

7.1. Reflections on planned next steps

For projects that involve listing and identifying available and publicly accessible datasets like that in **Kenya**, next steps / projects that can be useful are to invest more in skills development in:

- Data use / analysis
- Exploring how to transform the data into indicators that policymakers can use and leverage.

For projects which entail understanding the ecosystem of data producers and data needs like the one in **Sierra Leone**, next steps would be to engage with the government to make best use of the information included in the data ecosystem. This can include:

- Consolidation of data producers and of data
- Mounting portals to include some of the data.

Other potential future ideas include using this developed data ecosystem to start thinking about how to enhance coordination between the different data producers and systems and make them speak to each other in order to reach a consolidated, and eventually, a single source of information. The question is one of how to tie all these different sources together into a comprehensive education database warehouse. It is worth noting that the team would not have thought of this next step had they not known what the sources were initially. Thus, understanding the data ecosystem has been instrumental in shaping next steps.

The **Malawi** project aims to work on two research papers following data availability for the team, in collaboration with the ministry of education, the national examination board, and the national statistics office. In a second phase, there is a plan to make databases available for the research community to use. These databases would potentially include merged variables from other relevant databases so that researchers can find data on all the necessary variables for their study in one big database. Currently, this is the main bottleneck for the research community: researchers are unable to find data on all the variables they need for their studies in databases collected by the

government. There is also the issue of the time and effort required to clean and harmonise these datasets. This makes it more difficult for researchers to rely on these databases. Hence, many researchers prefer collecting their own data via primary data collection, which is inevitably undertaken on a small scale and therefore not representative, and thus not beneficial for policymaking. As expressed by a representative of the Malawi team, the aim is to:

> **"Unlock [...] those data by actually making it available for other users to use the same data to analyse something from a different perspective, maybe from the original person who collected it."**
>
> **– Malawi team representative, 2022**

In further phases, the project in Malawi also envisages:

- Reaching out for data on education that was collected by academia in universities.
- Conducting a cohort analysis using cohort tracking data that has never been utilised since its collection. This entails training officers within the ministry of education and the University of Malawi team on cohort analysis. The training will be delivered by an expert that CERT is currently looking for.
- Extending the scope beyond primary and secondary education levels to cover data on higher education.

7.2. Keep the collaboration going

In all data mapping projects, the next steps and dissemination strategies have to take into consideration how sensitive the findings / results are, and to keep the collaboration and trust chain with national entities going.

7.3. Pave the road towards data access

Data availability and accessibility is very important especially in contexts of low- and middle-income countries and in sub-Saharan Africa (SSA). In-depth data analysis is crucial for robust evidence that is necessary to inform policymaking. In-depth analysis of data can prevent inappropriate policy

decisions based on inaccurate research (↑Woolfrey, 2013), and enhance information governance.

In terms of data sharing, there is an important need to initiate the adoption of set procedures to follow when requesting data and to help establish clear policies and guidelines as to how to access data.

There is an assumption that development partners collect and create their own datasets in addition to those held by relevant ministries. Thus, in Malawi and Kenya, meeting and engaging with development partners who are involved in collecting those representative datasets (e.g., the World Bank, UNICEF) is another step that can be considered in paving the road towards making more data available and accessible.

7.4. Explore the new territory — data cataloguing

Data cataloguing is new territory for many countries in SSA. It not only involves investment in terms of funds, but also human capital investment in terms of technical skills, understanding of issues regarding confidentiality, ethics of use, and establishing user agreements. It is crucial to establish a common data platform or repository, or build on an existing one, to enable researchers and organisations to have access to education data as steps towards achieving quality education under SDG 4.

Bibliography

This bibliography is available digitally in our evidence library at https://docs.edtechhub.org/lib/KQUDWCXC

Adam, T., Agyapong, S., Asare, S., Heady, L., Wairimu, W., Mjomba, R., Mugo, J., Mukiria, F., & Munday, G. (2020). *Unlocking data to tell the story of education in Africa: webinar summary & synthesis*. Zenodo. https://doi.org/10.5281/ZENODO.4279156. Available from https://zenodo.org/record/4279156. Available under Creative Commons Attribution 4.0 International, Open Access. (details)

Hox, J. J., & Boeije, H. R. (2005). Data collection, primary vs. secondary. In *Encyclopedia of social measurement* (Vol. 1, pp. 593–599). Elsevier. http://www.joophox.net/publist/ESM_DCOL05.pdf. (details)

Lawson, L., & Heady, L. (2021). *Mapping education data in sub-Saharan Africa*. Education Sub Saharan Africa (ESSA). https://essa-africa.org/sites/default/files/inline-files/Draft%20Mapping%20Methodology_Unlocking%20Data.pdf. (details)

McBurnie, C., & Beoku-Betts, I. (2021). *Mapping the education data ecosystem in Sierra Leone*. EdTech Hub. (details)

McBurnie, C., Beoku-Betts, I., Waistell, D., & Nallo, M. (2021). *Advancing data-driven decision-making for school improvement: findings from the One Tablet Per School user testing programme in Sierra Leone*. MBSSE, EdTech Hub, Leh Wi Lan. https://doi.org/10.53832/edtechhub.0066. Available from https://docs.edtechhub.org/lib/MPRD5RKR. (details)

Ministry of Education. (2021). *Concept on mapping technical and vocational educational and training data in Kenya*. State Department for Vocational and Technical Training. (details)

Pambe, R., Kadzamira, E., Mjomba, R., Lawson, L., Onsomu, E., Adam, T., Heady, L., Mugo, J., & Tinto, T.-W. (2021). *Lessons learnt from education data mapping in Africa: Workshop summary and synthesis*. EdTech Hub. https://doi.org/10.53832/edtechhub.0052. Available from https://docs.edtechhub.org/lib/NNN9EZ7J. (details)

Rossiter, J. (2020). *Link it, open it, use it: changing how education data are used to generate ideas*. Center for Global Development. https://www.cgdev.org/publication/link-it-open-it-use-it-changing-how-education-data-are-used-generate-ideas. (details)

Siddiqui, N. (2019). Using secondary data in education research. *Social Research Update*, *68*. https://sru.soc.surrey.ac.uk/SRU68.pdf. (details)

Spencer, S. I. (2016). *Daughters of the Nile: Egyptian women changing their world*. Cambridge Scholars Publishing. Available from https://books.google.co.uk/books?id=oOmbDAEACAAJ. (details)

UNESCO. (2021). *Background information on education statistics in the UIS database*. UNESCO Institute of Statistics. http://data.uis.unesco.org/ModalHelp/OECD/background-information-education-statistics-uis-database-en.pdf. (details)

Unlocking Data. (2021, November). *Mapping education data in sub-Saharan Africa: Our experience from the Unlocking Data campaign and Sierra Leone*. https://docs.google.com/presentation/d/1WEUyqVEnxjx53eF-X4thxv8Gy2wWLvEdbN4aBuZiPbk. (details)

Vartanian, T. P. (2011). *Secondary data analysis*. Oxford University Press, USA. (details)

Woolfrey, L. (2013). *Leveraging data in African countries: Curating government microdata for research* (No. 22; DataFirst Technical Papers). DataFirst, University of Cape Town. https://www.researchgate.net/publication/281969484_Leveraging_data_in_African_countries_Curating_government_microdata_for_research. (details)