

FINAL REPORT ON SDGS DATA ECOSYSTEM MAPPING IN KENYA

June, 2019



TABLE OF CONTENTS

TABLE OF CONTENTS	2
LIST OF TABLES	6
LIST OF FIGURES	6
CASE STUDIES	6
1. EXECUTIVE SUMMARY	8
2. THE ASSIGNMENT	15
2.1 Background	15
2.2 Review Purpose and Objectives	16
3. METHODOLOGY	17
3.1 Overall Approach	18
3.2 Document Review	18
3.3 Stakeholder Interviews	18
3.4 The Goal Selection	19
3.5 Data Sources	19
3.6 Analytical Framework	20
3.7 Quality Assurance Mechanisms	21
3.8 Justification	22
3.9 Limitations of the Study	23
4. DISTINGUISHING OFFICIAL VERSUS NON-OFFICIAL DATA	24
4.1 Data for the SDGs in Kenya	25
4.2 Official Statistics	26
4.3 Non-official Statistics	27
5. FINDINGS AND DISCUSSION	32
5.1 Data Availability and Quality	33
5.1.1 Data Availability and Data Types	33
5.1.2 Data Collection and Validation	42
5.1.3 Barriers to Data Collection	46
5.2 Resourcing for Data Collection	47
5.3 Data Collaboration	48
5.4 Data Publication	50
5.5 Data Skills Gaps Across State and Non-State Actors	52
5.6 New Ways of Data Collection and Management	53
5.6.1 Use of Mobile Technologies	55

5.6.2 Use of Centralized Management Information Systems (MIS)	58
5.6.3 Use of Citizen Generated Data	59
5.6.4 Use of Big Data	61
5.7 Anonymization of Data	62
5.8 Politics of Data	63
5.9 KNBS Definitions, Guidelines, Norms and Standards	65
6. CONCLUSIONS	70
6.1 Data Availability and Quality	71
6.2 Data and Skills Gaps	73
6.3 New Ways of Data Collection and Management	74
6.4 Anonymization of Data	74
7. RECOMMENDATIONS	76
7.1 Data Availability and Quality	77
7.2 Data Publication	80
7.3 Data Skills Gaps	82
7.4 New Ways of Data Collection and Management	83
7.5 Data Collaboration	84
7.6 Politics of Data	84
8. SUMMARY	86
9. ANNEXES	88
9.1 Annex A: Interviews Conducted	89
9.2 Annex B: Some of the surveys undertaken by KNBS	91
9.3 Annex C: Fundamental Principles of Official Statistics (revised, 213)	92
9.4 Annex D: Data Collection Tool	93

Acronyms

ADP	Annual Development Plans
BMZ	German Federal Ministry for Economic Cooperation and Development
CABI	Centre for Agriculture and Biosciences International
CBOs	Community Based Organisations
CGD	Citizen Generated Data
CDR	Call Detail Record
CIDPs	County Integrated Development Plans
CIMES	County Integrated Monitoring and Evaluation System
CHVs	Community Health Volunteers
CoG	Council of Governors
CLTS	Community Led Total Sanitation
CPSB	County Public Service Board
CSOs	Civil Society Organisations
DHIS2	District Health Information System version 2
ECDE	Early Childhood Education in Kenya
FGD	Focus Group Discussions
GBV	Gender Based Violence
GCP	Gross County Product
GIZ	The Deutsche Gesellschaft für Internationale Zusammenarbeit
HIV	Human Immunodeficiency Virus
ICT	Information and Communication Technology
IFPRI	International Food Policy Research Institute
IGRTC	Intergovernmental Relations Technical Committee
iNGO	International Non Governmental Organization
KALRO	Kenya Agricultural and Livestock Research Organization
KEMRI	Kenya Medical Research Institute
KEPHIS	Kenya Plant health inspectorate Service
KEWASNET	Kenya Water and Sanitation Civil Society Network
KDHS	Kenya Demographic & Health Survey
KDSP	Kenya Devolution Support Programme
KIHBS	Kenya integrated Household and Budget Survey
KII	Key Informant Interviews
KNBS	Kenya National Bureau of Statistics
MCA	Member of the County Assembly
M&E	Monitoring and Evaluation
MDAs	Ministries, Departments and Agencies
MDGs	Millennium Development Goals
MIS	Management Information System
MoALF	Ministry of Agriculture Livestock and Fisheries
MoEST	Ministry of Education, Science and Technology
MoWS	Ministry of Water and Sanitation
MTPs	Medium Term Plans
NACOSTI	National Commission for Science, Technology and Innovation
NEMIS	National Education Management Information System
NGOs	Non Governmental Organizations
NIMES	National Integrated Monitoring and Evaluation System
NOFBI	National Optic Fibre Backbone

NSO	National Statistical Offices
NSS	National Statistical System
PDFs	Portable Document Format
PWDs	Persons with Disabilities
PMR	Program Monitoring Response
SDGs	Sustainable Development Goals
SGBV	Sexual Gender Based Violence
TB	Tuberculosis
TVET	Technical Vocational Education and Training
TWG	Technical Working Group
UNDP	United Nations Development Programme
WARIS	Water Regulation Information System
WASREB	Water Service Regulatory Board
WHO	World Health Organization
WSPs	Water Service Providers

LIST OF TABLES

TABLE 1: Official versus non-official statistics
TABLE 2: Data sources in Vihiga, Kisumu and Elgeyo Marakwet County Governments for the SDGs under study
TABLE 3: CSO Data sources for the SDGs under study
TABLE 4: KNBS detailed data sources for the SDGs under study
TABLE 5: Overview of SDG data sources
TABLE 6: Transferrable principles of data quality borrowed from the Data Quality Assessment Framework for Balance of Payments Statistics and the Principles of Fundamental Statistics.

LIST OF FIGURES

FIGURE 1: Important themes in this study. Data quality was a central theme
FIGURE 2: Producers of official and non-official statistics (Mkoswa, 2016)
FIGURE 3: Current data sources for the SDGs (with examples)
FIGURE 4: KNBS service charter
FIGURE 5: ICHIS Process/Data Flow.
FIGURE 6: Community Health summary data for MoH 516 from ICHIS
FIGURE 7: Decreasing levels of trust in data from Left to Right as per interviews
FIGURE 8: Data availability for the SDGs under study (applicable to both state and non-state actors)
FIGURE 9: Sample infographic: State of water, KEWASNET 2015/16 Annual Report (source is MoWS, KNBS)
FIGURE 10: Shifting from 'who' collects data to 'how' data is collected.
FIGURE 11: Proposed process for non-state data collection and validation
FIGURE 12: Recommendations on skills strengthening for different stakeholders for improved data quality.

CASE STUDIES

CASE STUDY 1: Data validation for Uwezo Project - 2011 to 2015
CASE STUDY 2: UNICEF Program Monitoring Response (PMR) - Kenya
CASE STUDY 3: Citizen Generated Data at scale. Uwezo Project influence in 2013
CASE STUDY 4: Global Fund Malaria Project (Source: AMREF Health Africa)
CASE STUDY 5: What happened to these Mega Projects in Kisumu?



1. EXECUTIVE SUMMARY

1. EXECUTIVE SUMMARY

The data mapping exercise focused on two main objectives:

- 1.** Conducting a data ecosystem mapping for Kenya: This includes data collected by official and non-official data producers, tools and methodologies used for data collection and the main data producers (and users) of the data. In order to allow for an in-depth analysis, the data ecosystem mapping has been limited to a number of selected SDGs that have been defined by key stakeholders, and focuses on data and processes at both the national and the county level (specifically Kisumu, Vihiga and Elgeyo-Marakwet Counties).
- 2.** Developing a set of recommendations on how to build an integrated multi-stakeholder approach to SDGs monitoring in Kenya based on the mapping exercise.

In terms of methodology, the approach entailed reviewing data collection, analysis, quality assurance processes and tools of both state and non-state actors at the national level as well as in three counties (Elgeyo Marakwet, Kisumu, Vihiga). The document review investigated existing legal frameworks, policies and guidelines around data collection, analysis and dissemination in the era of the SDGs. Overall, we considered the number of respondents a fairly representative sample for the mapping exercise given their level of expertise and seniority within participating organisations.

The study sought to investigate the actual practice in monitoring SDGs. The mapping exercise focused on five Sustainable Development Goals (SDG) namely, SDG 2: Zero Hunger, SDG 3: Good Health and Well-Being, SDG 4: Quality Education, SDG 6: Clean Water and Sanitation and SDG 7: Affordable Clean Energy - having been guided by prior meetings of key stakeholders. The mapping exercise relied on three main sources of data: secondary programme/project documentation/data, key informants, group interviews, and field observation. An analytical framework was designed to guide the content and methodology of the review; deliberately outlining the focus of the study, as outlined in section 3.6. The Quality Assurance mechanism incorporated a number of core principles to ensure a strong mapping exercise, these include inclusion, appropriateness, triangulation and transparency. Additional considerations/data sources and examples were incorporated to strengthen the justification while outlining the imitations of the study.

This report provides an overview of the SDGs data ecosystem in Kenya, focusing on data produced by selected county governments as well as civil society organisations. The objective is to find ways of fostering and strengthening collaboration between data producers on implementing SDGs in Kenya. The paragraphs below summarize the main findings.

Data sources: The mapping exercise confirmed that both state and non-state data producers collect and utilise data from both primary and secondary sources. In all counties, county governments and Ministries, Departments and Agencies (MDAs) collect data from primary sources such as using surveys and structured/unstructured interviews and observations as well as data from Management Information Systems (where they exist such as DHIS). Non-state actors employ more of qualitative methods (unstructured interviews, surveys, Focus Group Discussions, public forums and observations) of data collection as well as a lot of desktop research relying on existing Kenya National Bureau of Statistics (KNBS) and other MDA reports. The 2009 population census is one such critical example that is often referenced. Inherent however is the knowledge that there is a time-lag at any given time; such official statistics might be useful for a baseline - but not to inform operational level (or every-day) decisions, such as, the amount of drugs to be purchased at a health facility.

Key players: Kenya National Bureau of Statistics (KNBS) remains a key player in the data ecosystem as it acts as the producer of official statistics in Kenya. Generally, all the other MDAs, county governments and non-state actors regularly draw information from KNBS to inform programmes and plans, but they also collect their own data in specific areas. Non-state actors include CSOs (at national and subnational level), media, academia and private sector actors. For this study, the detailed list of interviewees and their organizations is provided in Annex A.

Data uses: What roles do data producers play? Data produced by county governments (namely Kisumu, Vihiga, Elgeyo Marakwet) mainly informs planning, policy formulation and budgeting; it also serves for engagement with the public during public participatory forums and educating communities. For example, Elgeyo Marakwet through the Public Participation Act (2014)¹, has instituted the Office of Public Participation which facilitates and coordinates public participation in the governance of the county. This office ensures that sessions are not constrained to traditional sectors, but that they cover the broad spectrum of public interests with clear context backed by data and evidence by not only facilitating budget process but any other issue of public importance - at least once in three months. Data produced by non-state actors is mostly used to highlight a social phenomenon (especially among minorities) or understand underlying causes in order to change (especially negative) situations and behaviours. Indeed, such studies (interpretivist research that includes purposive sampling techniques, mostly employed by non-state actors) are targeted at advocacy; but for it to inform policy, more rigor must be infused. This is evidenced in Twaweza's Uwezo project that

¹ <http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/ElgeyoMarakwetPublicParticipationAct2014.pdf>

has been monitoring numeracy and literacy scores for learners in East Africa since 2011 to ascertain if children (ages 6 to 16) are learning in class.

Data disaggregation: For the SDGs under study, statistics are mostly disaggregated by sex (both state and non-state actors), education levels, age (both state and non-state actors), location, and in some instances, disability (but mostly by non-state actors). The only KNBS surveys disaggregating by disability are population censuses and the Kenya Integrated Household and Budget Survey (KIHBS). Altogether, official statistics often lack disaggregation below the subcounty level. Disaggregation, especially by gender, enables us to understand trends and patterns but more importantly, develop evidence-based policies.

Resourcing for data: Finance and skilled personnel are key drivers for effective data collection processes. Human resources and personnel skills are important for successful data collection, analysis, dissemination and storage. However, inadequate financial resources are a key hindrance to fill existing data gaps and to regularly update data. Insights from non-state actors for example demonstrated that donor funding was inadequate and unsustainable in supporting the strengthening of data collection, management and use processes; this was characterized by KEWASNET who highlighted that for some years, they are unable to generate the CSO (Civil Society Organization) Annual Water and Sanitation Performance Report² due to lack of funds. Twaweza have core funding that has been aligned to their strategic objectives which allows them to undertake longitudinal research under Uwezo - building on previous years. Furthermore, due to inadequate financial resources most non-state actors do not possess in-house capability to employ dedicated staff for data collection and management, therefore having to rely heavily on consultants who did not leave data with the organisations. The mapping found there was inadequate infrastructure in some of the counties or among non-state actors to support effective data management and use within their organisations. When funding (which in many cases comes from external donors and which is not reliable) is no longer forthcoming or projects end, data collection may cease.

Challenges: Even though data is being collected, analysis and publication is not happening as regularly as it should. Reasons cited include lack of resources (financial and human), insufficient demand for the actual use of data for evidence-based decision and policy making (that is, demotivation because other criteria like politics are used to support decisions, not evidence). Other challenges include the inability to continuously monitor implementation (due

² This is a report generated by all the members of KEWASNET highlighting contributions by Kenyan CSOs in the water sector. Only 3 have been generated so far; this link is for the 3rd one (2015/16)
<http://www.kewasnet.co.ke/index.php/resources/kewasnet-publications/category/9-cso-reports?download=44:cso-report-for-print>

to understaffing and lack of efficient tools or technology) of programmes and projects, and lack of practical skills to analyze and interpret data. Another theme that came up is the issue of data illiteracy especially among citizens. The inability to understand the language of data adversely impacts civic engagement. Also, when data is collected by unskilled or untrained personnel, the quality of such data cannot be guaranteed.

Data validation: Methodological rigor is important if studies and research desire to inform decision and policy making. This study found that KNBS invites non-state actors to data validation workshops; but the opposite (that is, CSOs inviting KNBS to validate data results) does not happen. This contributes to a mistrust in data produced by non-state actors - especially that produced by the media. To remedy this, a 6 step process for the validation of data from non-official sources is prescribed in section 7.2 that has been developed after wide consultations with various data producers. There needs to be a shift from 'who' publishes the data to 'how' the data is published. A key insight is that documentation on the methodology will reduce mistrust in data among actors.

Key recommendations from the study on how to build an integrated multi-stakeholder approach to SDGs monitoring in Kenya are as follows:

Data Availability and Quality

- 1.** Increased collaborations between state and non-state actors, with KNBS being a focal point and taking lead, will generally ensure that duplication of resources expensed and interviewee fatigue during data collection are avoided. More importantly, data validation and methodology as prescribed by KNBS will be infused so that studies, programmes and research withstand the tests of reliability and validity as well as reducing of mistrust.
- 2.** Data education and data literacy is important - whether it is respondents understanding the important role of data collection and them responding to enumerators, or it has to do with citizens being able to interpret data. Data literacy is critical to ensuring that continuous supply of data feeds into planning and decision making for improved service delivery. Both state and non-state actors should drive data literacy.
- 3.** There should be distinct classification of data by KNBS:
 - a. Administrative data** - roads, disease burden, enrolment, does not need a lot of validity tests (it can be sampled randomly to ensure that it is representative of the population).
 - b. Survey data** - would require more rigor in terms of validity/reliability tests; KNBS should look at the research instruments; sampling using NASSEP V frame is very important for surveys.

Note that in this report, the use of official data and official statistics; non-official data and non-official statistics respectively; are used interchangeably throughout the report to mean the same thing.

Data Publication

- 1.** KNBS should develop data publication protocols in consultation with state actors (at national and county government level); these should be developed with and/or communicated with non-state actors.
- 2.** Also, non-state actors should improve on sharing/publishing findings with multiple stakeholders as they focus more on data visualization.
- 3.** For both state and non-state actors, increasing disaggregation to the lowest levels helps citizens relate to their own local targets, challenges and results so as to seek solutions to address them; as opposed to an overall comparison with national targets or results.
- 4.** KNBS should develop practical and operational guidelines that explain exactly how data is collected and verified before publication of data from non-state actors. Further, the county statistics bill should be fast tracked so as to empower counties to publish their own verified data.

Data Skills Gaps:

- 1.** Increased funding for data collection, analysis, dissemination and storage work by county governments, national government and donors will ensure that much needed data for programme decision-making can be sustained.
- 2.** Citizens should be given basic skills in data literacy. This also involves the use of infomediaries and community champions (such as teachers) to simplify findings, data and results; even when a study is targeting complex and abstract policy changes. Citizens need to understand data and policies therein that ultimately impact them.
- 3.** Locals are the people who best know what has been implemented (by state or non-state actors) or not. It is key for the success of devolution and for SDGs implementation.

New Ways of Data Management:

- 1.** The current methods of printed documents, storage of documents in flash disks, hard drives and other 'closed' media, within the county data ecosystem are not sustainable or

scalable. Having electronic data platforms and other back-up means will prevent loss of data and enable faster analysis.

2. Citizen Generated Data (CGD) : mechanisms and frameworks for institutionalizing CGD in the public sector will be required in order to support adoption and investments for proper data storage and analysis in order to leverage it for development.
3. Big data is currently not in use by any of the counties (or KNBS) except for Vihiga County which has set up a GIS lab that makes use of satellite imagery from NASA and Airbus; but plans are underway (they include creation of sound laws on manipulating personal mobile records) to incorporate mobile data sources into KNBS reporting.
4. Platforms that bring all stakeholder data making it interoperable are important because this helps to avoid duplication during program design, and data collection so that efforts complement each other to secure the missing data.
5. Stakeholders must strive to not just collect but also use the data - however little in decision making. Often politics, resources and other considerations take precedence over evidence.
6. Structured investments in infrastructure are a way through which County ICT infrastructure can be boosted in order to ensure that software and hardware required to support the data infrastructure is achieved.

Data Collaboration: Operating guidelines that coordinate data functions at the operational level but which still embody the original spirit of appreciating the distinctness, interdependence and respect for the mandate and institutions of all levels of government should be developed - possibly with coordination of CoG and IGRTC.

Politics of Data: Politics seem to play a role that undermines the use of data in implementation of planned interventions. This manifests as competing incentives, lack of good will, concerns around evidence and data and competing interests. It is therefore recommended that strong advocacy and holding different parties (the executive as well as the political arm of the county government) to account will ensure that politics do not supersede data and evidence in implementing policies.

Overall the report concludes that, there is a need for better collaboration and coordination between different data producers. Those working on the same SDGs have similar experiences and quite similar data; hence there are missed opportunities because of a lack of collaboration. Coordination will greatly help to ensure that data collected complements or updates existing data and to build trust among different stakeholders in the data ecosystem.



2. THE ASSIGNMENT

2. THE ASSIGNMENT

2.1 Background

In 2015, the world community reaffirmed its commitment to sustainable development. Through the 2030 Agenda, 193 member states pledged to ensure sustained and inclusive economic growth, social inclusion, and environmental protection, fostering peaceful, just, and inclusive societies through a new global partnership. The 2030 Agenda includes 17 Sustainable Development Goals, 169 targets and 244 indicators (of which there is general agreement on 232 indicators) all aimed at a universal integrated and transformative vision for a better world³.

To fully implement and monitor progress on the SDGs, decision makers need data that is accurate, timely, sufficiently disaggregated, relevant, accessible and easy to use. Whereas in Kenya, data availability and quality has improved over the years, statistical capacity still needs to be strengthened and data literacy enhanced at different levels. However, there are challenges with the national SDGs monitoring framework which include: fragmented data collection, data gaps in the official statistical system and limited collaboration between government, civil society and the private sector for data collection and usage.

Against this background, the German Federal Ministry for Economic Cooperation and Development (BMZ) commissioned the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) with implementing a component on SDGs monitoring and review in Kenya as part of the Strengthening Good Governance Programme. The SDGs component aims to build on the processes launched by the Government of Kenya and non-state actors to implement and monitor the SDGs and tackle the above-mentioned challenges. The component's objective is to strengthen the cooperation between governmental and non-governmental data communities that contribute to monitoring the implementation of SDGs in Kenya.

We understand that this will be achieved through three main areas of action:

- 1.** Catalysing dialogue and consensus-building through collaboration between government, civil society and the private sector;
- 2.** Establishment of norms and standards for (non-official) data collection and usage; and
- 3.** Piloting innovative data collection at the county level (in Vihiga, Kisumu, and Elgeyo-Marakwet).

³ <https://unstats.un.org/sdgs/indicators/indicators-list/>

2.2 Review Purpose and Objectives

Through its SDGs component, the GIZ Good Governance Programme commissioned a mapping of the SDGs data ecosystem in Kenya. The objective of this assignment is two-fold:

- 1.** Conduct a data ecosystem mapping for Kenya. This will include data collected by official and non-official data producers, tools and methodologies used for data collection and the main data producers (and users) of the data. In order to allow for an in-depth analysis, this data ecosystem mapping will be limited to a number of selected SDGs indicators to be defined by key stakeholders. It will focus both on the national and on the county level (specifically Kisumu, Vihiga and Elgeyo-Marakwet Counties).
- 2.** Develop recommendations on how to build an integrated multi-stakeholder approach to SDGs monitoring in Kenya based on the mapping exercise.



3. METHODOLOGY

3. METHODOLOGY

3.1 Overall Approach

The study reviewed data collection, analysis, quality assurance processes and tools of both state and non-state actors at the national level as well as in three counties (Elgeyo Marakwet, Kisumu, Vihiga). It also identified successes and challenges related to tracking of SDGs based on the national SDG Framework Implementation⁴.

3.2 Document Review

The document review section investigates existing legal frameworks, policies and guidelines around data collection, analysis and dissemination in the era of the SDGs. Generally the document review includes: Journal articles examining various issues impacting NSSs, Kenya Annual Statistical Abstracts (2016, 2017 and 2018), County Statistical Abstracts of 2015⁵, UNSTATS SDG Indicator Metadata Repository, legal provisions in the Constitution of Kenya 2010, the County Governments Act 2012, the Statistics Act of 2006, as well as policy briefs from different members of the SDG Kenya Forum such as KEWASNET, AMREF, Twaweza. Other documents reviewed include the County Integrated Development Plans for FY 2018/2019 and Annual Work Plans. Lastly, important documents on guidance by the Kenya National Bureau of Statistics (KNBS) on collecting, analysing and publishing data were reviewed.

3.3 Stakeholder Interviews

The list includes county officials (Departmental Chief Officers or Directors), Head of ICT, and other user department staff in economic finance and planning, health, agriculture, water, and external stakeholders (citizens/Project Management Committee members, and CSOs, iNGOs) working in the defined SDGs producing (and consuming) data. Apart from being subjects of informant interviews, staff from the KNBS were also consulted for quality assurance during stages of the process to ensure methodological rigor of the tools (used in data collection). As stated, in developing the SDG indicator monitoring framework, KNBS considered the country priorities as well as ease of data collection. The full list of stakeholder interviews is provided in **Annex A**.

⁴ This is for the implementation of the Agenda 2030 for Sustainable Development in Kenya, 2017

⁵ <https://www.knbs.or.ke/county-statistical-abstracts/>

3.4 The Goal Selection

At the national level, monitoring of SDG progress ought to be through the National Integrated Monitoring and Evaluation System (NIMES). At sub national level, tracking progress outlines that each CIDP should be undertaken through the County Integrated Monitoring and Evaluation System (CIMES). Interviews and observations at national and subnational were undertaken to investigate the actual practices and implementation of both frameworks on the ground (a gap exists, see section 6.3).

The goals under consideration were:

- 1. Goal 2:** End hunger, achieve food security and improved nutrition and promote sustainable agriculture
- 2. Goal 3:** Ensure healthy lives and promote well-being for all at all ages
- 3. Goal 4:** Ensure inclusive and equitable quality education and promote lifelong opportunities for all
- 4. Goal 6:** Ensure availability and sustainable management of water and sanitation for all
- 5. Goal 7:** Ensure access to affordable, reliable, sustainable and modern energy for all

The goals were selected after lengthy discussions amongst the key stakeholders (see section 3.3 and Figure 3 for a list of those supplying data)

3.5 Data Sources

The study relied on three main sources of data: secondary programme/project documentation/data, key informants, group interviews, and field observation.

- 1. Secondary sources:** Reviewing programme/project and other relevant policy guidelines and reports (from state and non-state actors), documents and data has allowed the project team to gain a fuller understanding of vital sources of data available and the important role they can/are playing. This study also checked how existing datasets (from the various organizations as well as county government departments) resonated with the local county's challenges and priorities (especially the sectoral short term ones).
- 2. Key Informants (individual and group):** Semi-structured interviews with a selection of stakeholders are an ideal method of obtaining in-depth, qualitative information. We interviewed a selection of key informants (sometimes individually or in a group discussion format) at the county level and some national level actors. The interviews were used to obtain views of informants (see **Annex A** for the list). The survey was intended to provide a quick snapshot of the knowledge and practices of respondents across the issues covered by the 5 SDGs in their respective contexts.

-
- 3. Field Observation:** Field visits to observe practice on the ground. The notes taken during the field visits have been incorporated into the findings discussion in the final report in order to triangulate and validate views from the multiple actors at the county level.

3.6 Analytical Framework

The analytical framework guides the content and methodology of the review; that is to outline what the study should look at in order to answer the main questions in the Terms of Reference. With the Terms of Reference as a start, the analytical framework has guided data gathering and discussions with the various stakeholders – aiming to answer the following questions:

- 1. Do you involve National Statistical Systems in the development of your methodology, collection of data, analysis, or data dissemination?**
 - a. What policy or legal frameworks and guidelines do you take into consideration?
- 2. What SDGs does your work monitor?**
- 3. Who are other entities you work with in your projects**
 - a. Name the organizations
 - b. What role do they play?
 - c. How are they or the partnerships funded?
- 4. What is your methodology?**
 - a. How do you ascertain validity and reliability of findings?
- 5. How do you:**
 - a. Collect and store data?
 - b. What is the frequency of data collection?
 - c. Analyze the data?
 - d. Share findings?
 - e. Data uses?
 - f. Protect data?
- 6. If the indicator involves multiple components from different data sources, please describe how each individual component of the indicator is collected**

The diagram below outlines the main themes covered by the data collection tool (data quality – timeliness and availability are both important dimensions of quality) – which is shared in the **Annex D**.



Figure 1: Important themes in this study. Data quality was a central theme.

3.7 Quality Assurance Mechanisms

The project team incorporated the following quality assurance mechanisms in our initial approach and throughout the data collection and analysis process:

- 1 Inclusion:** the project team’s approach to data collection utilised a transparent, participatory, consultative process that meaningfully involves and engages key (state and non-state) stakeholders from the 5 SDGs. We sought to document a spectrum of experiences from data collection and analysis to the dissemination process. We also specifically investigated what collaborations are purposely put in place by various actors with KNBS when they are programming or collecting data.
- 2 Appropriateness:** we have checked the relevance of our tool and the participants responses to ensure that the data produced answered the key questions in the Terms of Reference.
- 3 Triangulation:** The project team worked to ensure that appropriate sources of evidence for undertaking the review were used in a technically appropriate manner. The project team collected and analysed data from different available sources in order to triangulate information – checking and corroborating findings from state and non-state stakeholders to ensure that they are consistent and accurate with what is on the ground so that requisite

recommendations can be made.

- 4 Transparency:** In order to ensure transparency, the project team shared the data collection tool with representatives of the stakeholders (at national and subnational level) working on the ground on the 5 SDGs as well as with KNBS for validation. Further, the project team has conducted a report validation exercise with multiple stakeholders as an important platform to discuss initial findings, gather feedback, and identify any errors in fact or misinterpretations. Additionally, where permission is granted, the project team may undertake additional Key Informant Interviews (KIIs) and Focus Group Discussions (FGDs) for later content analysis and provide clarification in the event of data discrepancies.

3.8 Justification

The purpose for which data will be used has a lot of influence on the data collection methodology, and the conclusions drawn from this data; that is, the problem which the data tries to address, for example, policy, advocacy, economic and development planning, amongst others.

But increasingly, with the implementation of the SDGs, there is an emergence of non-traditional indicators (for example Indicator 7.3.1 is energy intensity measured in terms of primary energy and GDP) for which NSO and MDAs might not have data. Also, there are cases where data that is disaggregated by location is simply not collected; for example, due to methodology constraints, KNBS does not collect data on child mortality lower than the national level⁶. What happens in the absence of official data? In most cases, non-official statistics have been used to make plans and decisions⁷; thus the infusing of sound methodology would greatly improve the quality of decision making.

Additional considerations/data sources and examples to be examined by the study include:

- 1.** Some organizations are already publishing data that influences policy. Twaweza⁸ is an NGO that is collaborating closely with KNBS on the Uwezo Initiative which sought to answer the most basic of questions: Are East African schools succeeding in the responsibility of ensuring children are learning? This mapping shall evaluate the process and methodologies that have been adopted for projects like these which have informed policy making.

⁶ Prior to 2013, this data was not disaggregated since counties did not exist. Presently, this data can be found in DHIS2 at facility level; child deaths are also reported to chiefs (that is, those not reported to facilities) under the office of the president. Surveys cannot adequately capture this information. This then results in aggregation at the national level.

⁷ *ibid*

⁸ <https://www.twaweza.org/>

2. Citizen-generated data has the potential to enrich existing data by complementing gaps that are missing. For instance, citizen groups could want to ensure the reliability and accuracy of their data by comparing it against official data collections or prediction models.
3. Big Data has been described as data sets that are impossible to store and process using common software tools, regardless of the computing power or the physical storage at hand. It will be important to investigate if any of the stakeholders are using big data and how they are using it.

3.9 Limitations of the Study

The study has been limited to the 5 SDGs mentioned in Section 3.4 and 3 counties (Vihiga, Kisumu and Elgeyo Marakwet County). The selection of the counties aimed to maximize the data processes that exist in each of these counties both from a state and non-state perspective. The terms official data and official statistics and non-official data and non-official statistics are used interchangeably in this report where the term statistics or data has been used. Also, the study distinguishes between KNBS, state actors (county governments, MDAs) and non-state actors (to represent CSOs, media, academia, private sector).

Unfortunately, time allocation and scope of this study did not allow gathering of actual data sets for mapping with SDG indicators; however, the study did find that a plethora of data is generated by MDAs, county governments and non-state actors. Most of such data as it was realized in this study resides behind 'closed' systems and tools. Consequently, linkages between the available data and the specific SDG indicators could not be exhaustively achieved given that indicator mapping is the most accurate way to tie in CIDP indicators and the SDGs indicators. This is an interesting area for further exploration.

Despite this limitation, all the information gathered during this study (including data from KIIs) was thematically analysed in line with the study objectives and summarized tables are provided in Section 5.1.1 of the report.



4. DISTINGUISHING OFFICIAL VERSUS NON-OFFICIAL DATA

4. DISTINGUISHING OFFICIAL VERSUS NON-OFFICIAL DATA

The following section has been introduced (separately) in order to provide a detailed understanding of official and non-official statistics and the respective producers.

4.1 Data for the SDGs in Kenya

Lessons learnt from the MDGs implementation in Kenya indicate that the objectives of the 2030 Agenda cannot be achieved without a strong coordinated effort from multiple stakeholders drawn from both state and non-state actors. In addition, it is understood that the Kenya SDG Roadmap⁹ will be implemented at both national and subnational levels through the Medium-Term Plans (MTPs) and respective County Integrated Development Plans (CIDPs).

With 232 indicators to be monitored, the SDGs provide a monumental challenge for countries and development partners alike, who are under pressure to be more efficient and cost-effective in providing data in an increasingly overstretched and under-resourced environment. In Kenya, the Kenya National Bureau of Statistics (KNBS) has mapped out 128 indicators for which data will be readily available within a short time (work is going on to increase the number of indicators¹⁰) as well as the country priorities regarding SDG monitoring. Among other things, this includes improved integration of different data sources, both old and new, and the implementation of new and improved methods of data collection. At the national level, monitoring of SDG progress should be done through the National Integrated Monitoring and Evaluation System (NIMES). At sub national level, tracking progress outlined in each CIDP should be undertaken through the County Integrated Monitoring and Evaluation System (CIMES).

However, in implementing the Kenya SDG Roadmap, certain specific data challenges have been identified, that impact monitoring of progress as follows:

- 1.** Lack of disaggregated data since often data (such as that on Food Production – Goal 2 on Zero Hunger) will only be aggregated to the sub-county level; Goal 6 data on clean water and sanitation is aggregated to the county level.
- 2.** The frequency of data collection is inadequate for operations planning.
- 3.** Lack of baseline data for some of the SDG targets (for example, Goal 7 on affordable clean energy; Goal 14 on marine life – though not under review in this study)

⁹ <http://planning.go.ke/wp-content/uploads/2018/04/SDG-Implementation-plan-2030.pdf>

¹⁰ Kenya SDG Implementation Plan (2017)

-
4. Some data simply does not exist; or is not traditionally collected by KNBS (for example Goal 7 on affordable clean energy).
 5. The road map has recognized that there will be a need for capacity building of the National Statistical System to ensure data availability. Incorporation of non-state actors in the NSS is a specific strategy cited in the road map¹¹ so as to trigger new data collection tools and technologies.

In this study, it was found that for the sectors which are devolved at subnational level, county governments are more proactive since the data directly informs planning activities. For Goal 2, Goal 3, there is better and more regular data than was previously the case before devolution (though KNBS gets data from other state corporations such as Sugar Board, Agricultural Finance Corporation, among others). For Goal 4, counties have richer and more accurate data for ECDE and TVET since these are the devolved functions. There are tools and expertise to collect data on Goal 6 but most of this data is not disaggregated. Lastly, on Goal 7, we are not collecting enough data - especially at the subnational level (see summary in section 5.1).

4.2 Official Statistics

Official statistics are statistics produced and published by governments or government agencies; and which must be approved by a person of a certain designation in the government echelon, such as a head of the NSO. Official statistics are a public good. The NSO in Kenya is the KNBS. Besides primary data collection through various surveys (see Annex B) most of the available data originates from sectoral MDAs and county governments which collect the data.

The term “official statistics” is often also extended to data ‘produced’ by UN agencies and other multilateral organizations; but again a scan of the UNSTATS SDG Indicator Metadata Repository¹² reveals that the UN and multilateral institutions draw data (mostly aggregated at national level) from in-country NSOs and from international bodies such as the World Bank, the International Energy Agency, UNICEF, the World Health Organization, the World Food Program, among others.

Non-state actors often reference official data sources. For example, Kenya Water and Sanitation Civil Society Network’s (KEWASNET), Kenya’s key WASH statistics, cites data

¹¹ Ibid,

¹² <https://unstats.un.org/sdgs/metadata/>

sources such as The United Nations Children's Fund (UNICEF), World Health Organization (WHO), and the Ministry of Water and Sanitation (MoWS), among others¹³.

In this study, the terms “official data” and “official statistics” are used interchangeably - since the definition offered for official statistics is the same one offered for official data.

4.3 Non-official Statistics

Progress in data collection and innovative applications has led to a radical shift in how data and information are produced and utilized for development policy and decision-making. Technology has harnessed new data sources and new data users on the statistics platform hitherto unknown to or unfamiliar with traditional official statistics users. Today, NSOs are faced with the challenge of managing the exponential growth of non-official data, increased demand for more timely data, as well as the growing need for data from non-traditional sources for emerging sectors like affordable clean energy (SDG 7).

Non-official statistics are produced by non-government organisations and/or departments, such as independent research bodies, academic research, civil society organisations, commercial and market research organisations, the private sector and citizens themselves. At times they are produced by government departments but for internal use. It ranges from global surveys (for example, Transparency International’s Global Corruption Barometer¹⁴) to personal, qualitative data generated by some of the world’s most marginalised people at the local level (e.g. citizen-generated data). It also includes data translated from publicly available (open) data sets to track a specific phenomenon/issue (Transparency Index) and information collated through expert assessments (CIVICUS’ annual State of Civil Society Report). Another example encountered during field work was in Seme Central Ward where the MCA uses a sign-up book at her office as a data collection tool. Residents who visit her office for assistance or report incidences sign up and register information such as name, age, contact information, location and the issue that one is reporting. This has been the ‘only’ way for her to create disaggregated data. With this information she is able to know what issues are experienced by citizens in her administrative area and plan for possible interventions. This is a unique case, in that whilst she does not fall within the traditional taxonomy of a data producer, the practical value to improved decision making cannot be overlooked.

In summary, all of these data sources play a critical role in measuring progress against the SDGs and the 2030 Agenda.

¹³ <http://www.kewasnet.co.ke/index.php/resources/kewasnet-publications#>

¹⁴ https://www.transparency.org/news/feature/global_corruption_barometer_citizens_voices_from_around_the_world

Table 2: Official versus non-official statistics (See source¹⁵)

OFFICIAL STATISTICS	NON-OFFICIAL STATISTICS
<ol style="list-style-type: none"> 1. Provide quantitative and/or qualitative information on all major areas of citizens' lives, such as economic and social development, living conditions, health, education, and the environment. 2. This data is nationally representative, because it is obtained from complete censuses or very large-scale national sample surveys and conforms to international definitions and classifications. 	<ol style="list-style-type: none"> 1. The publication of the data does not need any clearance of the government such as official statistics. 2. Organisations producing non-official statistics have the liberty to publish them at will.
Merits	
<ol style="list-style-type: none"> 1. Statistics are produced with certain periodicity allowing a possibility of developing trends over time; 2. They allow comparison among nations; 3. Offices responsible for production and publication have legal mandate in their respective countries; 4. Data from censuses are good benchmarks for sample surveys and studies carried out in the respective countries in subsequent periods; 5. They are used in assessing sectoral performance of the socio-economic aspects of the country from time to time; and 6. They enjoy international backing and guidance such as from the United Nation Statistics Division (UNSD). 	<ol style="list-style-type: none"> 1. They can be very detailed and if properly collected and analysed, can be equally good sources as census data; 2. They can be produced cost-effectively since they are/can be obtained as by-products of administrative functions within the institutions; 3. They respond fast to contemporary data needs to address issues of interest to the population in all socioeconomic spheres; 4. They address almost all socio-economic areas; and 5. In some fields, especially in scientific ones, they are the only source of such statistics.
Demerits	

¹⁵ ibid

1. Overall, they are not disaggregated (the county, and in some sectors like agriculture, the sub county is the lowest level; health and education data is sometimes shared at the ward level¹⁶);
2. They do not portray the current situation on the ground due to periodicity of collection; lagging behind needs¹⁷;
3. Though not applicable for this study, Mkoswa (2016) cites other challenges which NSS's globally are facing:
 - a. Lack of regular funding leads to sample sizes that are not statistically sound or infrequent undertaking of vital surveys.
 - b. Interference may lead to the production and/or publication of partial statistics¹⁸;
 - c. Production and publication of these statistics are biased to economic aspects. There are enormous gaps in social statistics which have direct impact to people's lives¹⁹.

1. At times sample sizes used are too small to warrant meaningful inference to the whole population. This means they have, at best, to be taken as case studies, never to be used to generalise to the parent population;
2. Improper handling of statistics related information by non-statisticians in institutions leads to the production of unreliable statistics;
3. Lack of political will and low statistical literacy prevents harnessing potential in filling data gaps in official statistics;
4. Most non-official statistics lack important supporting background information/metadata such as the sample sizes, sizes of population of interest, sample units' selection methods, data collection methods, estimation methods, and data analysis procedures.

¹⁶ Evidenced by County Statistical Abstracts from KNBS for Vihiga, Kisumu and Elgeyo Marakwet for 2015 and Annual Statistical Abstract (2018); As well as a view of data on the Kenya Open Data Portal. Education: <http://www.opendata.go.ke/search?q=education&tags=ministry%20of%20education>; Health: <http://www.opendata.go.ke/search?q=health>

¹⁷ Evidenced by over-reliance on KNBS data mostly for baseline data across the board.

¹⁸ For example, in 2018, the KNBS presented a bill before the National Assembly to have laws changed to allow the Director General to nullify fabricated census results. It is believed this happened in Garissa, Mandera and Wajir in the 2009 housing and population census.

Psirmoi, D. (2018). *KNBS boss wants powers to revise, nullify census results*.

<https://www.standardmedia.co.ke/article/2001287256/knbs-boss-wants-powers-to-revise-nullify-census-results> (viewed 4/06/2019)

¹⁹ Many country NSS's focus and invest on economic aspects and less on social statistics which are also very important for planning.

Advantages of incorporating alternative data sources into monitoring SDGs progress are shared below:

- 1. Filling data and capacity gaps:** Non-official data can complement official sources of data, fill data gaps that exist in a timely way and supplement official reporting when current and up to date data is not available. In particular, civil society and citizen-generated data can help provide an accurate snapshot of SDGs implementation progress in local contexts, including amplifying the perceptions and voices of those typically marginalised and hard to reach.
- 2. Driving innovation and building capacity:** Having a dynamic data ecosystem can help drive innovation by bringing together a diverse range of actors working together as part of data production and usage. Collaborative working between NSOs and non-official data producers can also help develop new methods, build one another's capacities, skills and shared practices.
- 3. Broad ownership:** Inclusive data production will also foster data ownership across society and help make public engagement in the SDGs agenda more meaningful. Non-official data can also help improve the accuracy and impartiality of official reporting, and raise the alarm if these processes become politicized. This plays a key role in ensuring the legitimacy of our collective data, and painting a truly accurate picture of progress towards the SDGs.
- 4. Accountability:** Policymakers not only need data to make decisions, but civil society, opposition politicians, activists and the media need it to hold all stakeholders to account. Non-official data offers a crucial check and balance that can help ensure that official data portrays the genuine reality within society.
- 5. Improved decision making:** Development decisions and priorities cannot accurately be made with incomplete/missing or poor-quality data that is outdated and unreliable. Without data and meaningful information citizens cannot meaningfully participate in public discourse on projects, budgets, among others. Consequently, development activities cannot be properly monitored or the implementers held accountable.

See summarized diagram below of data producers.

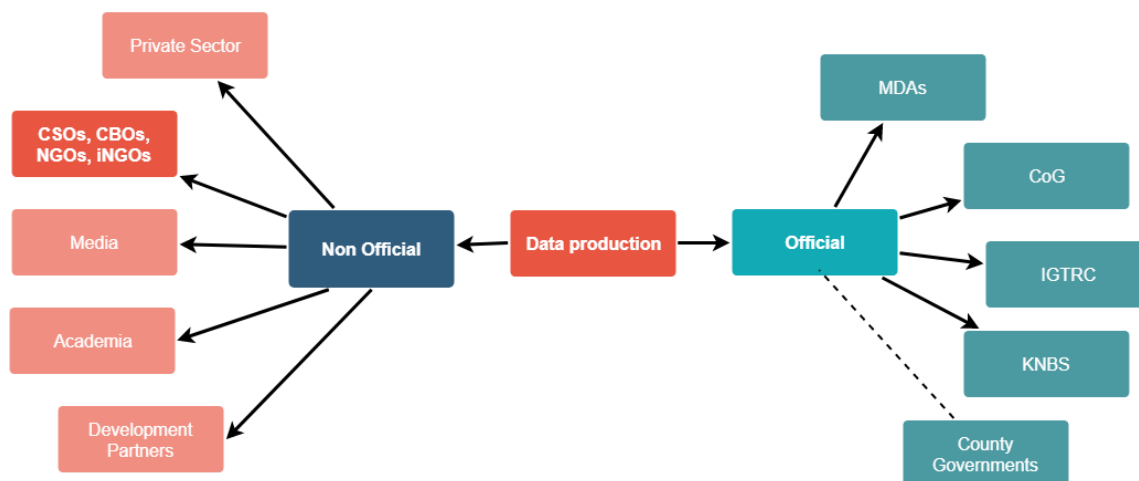


Figure 2: Producers of official and non-official statistics (Mkoswa, 2016)²⁰

²⁰ Mkoswa, Z.E. (2016). Beyond Official Statistics. International Journal of Applied Science and Technology. 6(2), pp 100-107



5. FINDINGS AND DISCUSSION

5. FINDINGS AND DISCUSSION

5.1 Data Availability and Quality

5.1.1 Data Availability and Data Types

a. County Government Data Sources

County governments are producing data which is being cited in the County Statistical Abstracts and Annual KNBS abstract (as will be seen in section 5.3 on data collaboration).

Results: Data sources
<p>All counties under study: Apart from registers that contain data on health and education the most common sources of data for county governments are questionnaires, Key Informant Interviews (KII), Focus Group Discussions and Public Participation Forums. In rare cases, secondary sources of data are also used because data from secondary sources (other MDAs such as Kenya Agricultural and Livestock Research Organization (KALRO), Kenya Medical Research Institute (KEMRI), Kenya Plant health inspectorate Service (KEPHIS), Kenya Bureau of Standards (KEBS), KNBS, Kenya Meteorological Department, among others) are vital but are not necessarily created by departments at the county level.</p> <p>Generally, county staff (agriculture extension officers, engineers managing water facilities contractors, staff in education department as well as Departments of Finance and Economic Planning) are all involved in collecting sectoral data from these sources.</p> <p>See Table 4 for details.</p>

Table 2: Data sources in Vihiga, Kisumu and Elgeyo Marakwet County Governments for the SDGs under study

Organisation / Department	Where is data published	Data source and how data is collected	Frequency of collection	Level of disaggregation	Related SDGs and/or SDG targets
MoEST MDAs and County Department of Education	KIHBS on KNBS Website	Surveys at household level. Indicators: school attendance, highest education level, School feeding programme, Literacy	5 years	County, Age, Sex, Education Level, Type of school (day/boardings)	Target 4.1, 4.2, 4.3, 4.4, 4.6,

KNBS, MoALF MDAs and County Department of Agriculture, Livestock and Fisheries	National Crop Conditio ns Bulletin	Weather outlook, crop conditions, market information	Monthly	By region and county By crop	Target 2.4
KNBS, MoE	KIHBS on KNBS Website	Main source of lighting, cooking fuel	5 years	By type of fuel	Cannot match
KNBS, MoWS	KIHBS on KNBS Website	By source of the water, access to drinking water, methods to make water safe, time to fetch water, volume of water used	5 years	County	Target 6.1
KNBS, MoH, County Department of Health	KIHBS, KDHS ²¹ on KNBS Website	Morbidity, Health insurance cover, child health, Place of delivery, Assistance during delivery, immunization, diarrhoea	5 Years	County, gender, age, employee/employ er, age	Target 3.1, 3.2, 3.4

Some important notes are made here; the mere existence of data should not be construed to mean that data is available. Availability of data has to do with it not residing in individual computers, laptops, phones, siloed departments but rather that it is relatively easy to access and download (both internally to the county government and externally). Also, the availability of data should not be assumed to extend to all the stages of research generation (sampling, frequency, data collection, analysis) and dissemination equally. Often, data may exist but it is also not published (so not available). Similarly, KNBS also technically argue that administrative

²¹ 2014 Kenya Demographic And Health Survey.
<https://www.knbs.or.ke/download/2014-kenya-demographic-and-health-survey/>

data that may be drawn from a Management Information System (MIS) might not necessarily pass the tests of random sampling. For example, there are children who do not report to school but who are of school going age; there are patients who do not seek treatment from health facilities. These are important ‘pockets’ of population and insight missed if the right sampling frame is not utilized (hence the use of the NASSEP V frame and production of the KIHBS is a remedy to this problem).

In addition, this study finds that where there is a centralized Management Information System (MIS), more frequent and hence higher quality data is available. Most of the data available for monitoring indicators is quantitative, and as that, measuring outputs. Outcome and impact indicators level data - though present (for example ECDE enrolment rates - is not abundant - as reflected in the second generation (2018-2022) CIDPs. Not many departments (at the county level) have clear baseline data and therefore mainly depend on KNBS data produced through KIHBS and other surveys. But for operational decision making, the county government departments do more regular data collection for performance monitoring; the frequency of collection can be monthly (for example health indicators), quarterly (data on infrastructure) or annually (for example agriculture data on crops, livestock).

An additional note is provided here on data use; how data will be used impacts how and why it is collected. At the county level, common uses of data include planning and formulation of development interventions by county government departments. The data is also required by different committees of the County Assembly. Thus, data is crucial in influencing budget processes and ensuring that resources are allocated according to needs as well as to highlight which areas and sectors need urgent attention²². Data is also used for community engagement, especially as a reference point when determining which projects should be prioritised for county Annual Development Plans (ADP). In all the counties under study, it was reported that citizens are becoming more assertive and ask a lot of questions during public participation meetings. For example, in Elgeyo Marakwet, sometimes county government proposals are rejected during the participatory budgeting process when they do not meet the needs and priorities of the community. With increasing and sometimes competing needs placed upon county governments, prioritising and phasing of different projects becomes crucial. Availability of data is thus important in making evidence-based decisions and engaging citizens during public participation forums. Finally, data is also used to monitor progress.

b. Non-state Actors Data Sources

²² Note however, it was noted by county government staff that more can be done to educate citizens on the complex budget processes and what is expected of them during public participation. CSOs need to plug into this deficit.

In Kisumu (and much similar to Vihiga and Elgeyo Marakwet), CSOs are involved in data collection on housing, water, people living with disabilities, education and gender. Being organisations that work closely with citizens at the county level, primary data collected remains important in advocating for different development and service interventions by the government, for citizens. However, as far as getting baseline information, all CSOs interviewed indicated they rely on secondary data for their baseline, especially from the county government, national government and from KNBS.

Some CSOs (like Twaweza²³, CABI²⁴) are able to sustain regular collection of data - either to monitor outcome and impact level indicators or to provide valuable data services. But as has been intimated, the continuous collection of data for programmes depends on availability of funds.

Table 3: CSO Data sources for the SDGs under study

Organisation / Department	Where is data published	Data source and how data is collected	Frequency of collection	Level of disaggregation	Related SDGs and/or SDG targets
TWAVEZA	Uwezo Annual Assessments on website	Numeracy and literacy skills, teacher attendance, impact of socioeconomic factors like poverty on academics	Annual from 2010 to 2016	Country, Regional	Target 4.1, 4.2
KEWASNET	Annual Report on website	Survey conducted on member CSOs. Main focus is safe water coverage	Annual	County, By Region	Target 6.1
AMREF	Annual Reports, Project specific	Human resource for health, innovative service delivery,	Annual	County, project based.	Target 3.1, 3.8, 3.c

²³ Data is disaggregated by gender, type of school attended, urban/rural and location information (county, subcounty, ward), age (they assess ages 6 to 16), school grade and number of children in school or out of school.

Household level data is disaggregated by the gender of household head, region, level of education of the household head, as well as level of education of the matriarch of the house.

²⁴ See the Plantwise Knowledge Bank: <https://www.plantwise.org/KnowledgeBank>

	reports on website	healthcare financing solutions. Please note that AMREF also runs a health facility thus they are collecting data on many more indicators.			
Access Coalition	No data	Access to clean energy	No data	No data	Cannot match
CABI	Projects with dedicated online tools. Plantwise Africa Soil, Health Consortium, Invasive Species on CABI website	Plant health, invasive species monitoring, soil health, among others	Project specific	Global	Target 2.3, 2.4

c. KNBS Data Sources

This section is provided because an understanding of where official data is sourced also benefits non-state actors; it also practically informs the broader data availability question for the SDGs.

Generally, KNBS not only implements their own surveys (see Annex B), but also complements their data by sourcing data from other ministries. KNBS collects data from ministries that is related to: 1) expenditure related information by the relevant MDA or county from the national treasury, and 2) public service provision by sectoral MDAs or counties and achievements on various indicators that are of interest.

KNBS collects expenditure information from the national treasury as this is used to inform the economic surveys. Some of the disaggregation for data collected above include sex, age, region (location), among others. Another key product developed in 2019 is the Gross County Product²⁵ (GCP, 2019); a breakdown of Kenya’s Gross Domestic Product (GDP) that estimates the current size, structure (and potential) of the economy of all 47 counties. This was informed by existing surveys and data sets collected from the county government departments.

²⁵ <https://www.knbs.or.ke/download/gross-county-product-2019/>

Table 4: KNBS detailed data sources for the SDGs under study

Organisation / Department	Publication of data	Data collection methodology (e.g. survey, FGD)	Frequency of collection	Level of disaggregation	Related SDGs and/or SDG targets
KNBS	Gross County Product on KNBS Website	Surveys on income, production and expenditure in all 47 counties KNBS Surveys and censuses ²⁶	Annual	County	Target 2.3
KNBS	Malaria Indicator Survey on KNBS Website	Determining the progress of key malaria intervention, assess malaria parasite and anaemia prevalence among children aged 6 months to 14 years	5 years	Community level, age, four epidemiological zones in the country, gender	Target 3.3
KNBS	Kenya AIDS Indicator Survey on KNBS Website	One off survey commissioned in 2017. The data provides information needed for advocacy and for planning appropriate interventions for HIV prevention, treatment and care	Last ²⁷ produced in 2017	Gender, age	Target 3.3

²⁶ The Kenya Population and Housing Census of 2009 is the most important of all censuses and surveys

²⁷ Last production date means it is not a regular survey.

KNBS	Agriculture and Livestock Survey on, KIHBS on KNBS Website	Crops and Livestock statistics on average price	Last produced in 2012	National	Target 2.c
	County statistical Abstract. For example, statistics on slaughtered livestock or food security	Livestock population census	Last produced with Population census of 2009		

d. Discussion

Different organisations (at the county level – both county governments as well as CSOs and CBOs) collect and produce data that is relevant to their respective areas of expertise as well as programmes and projects.

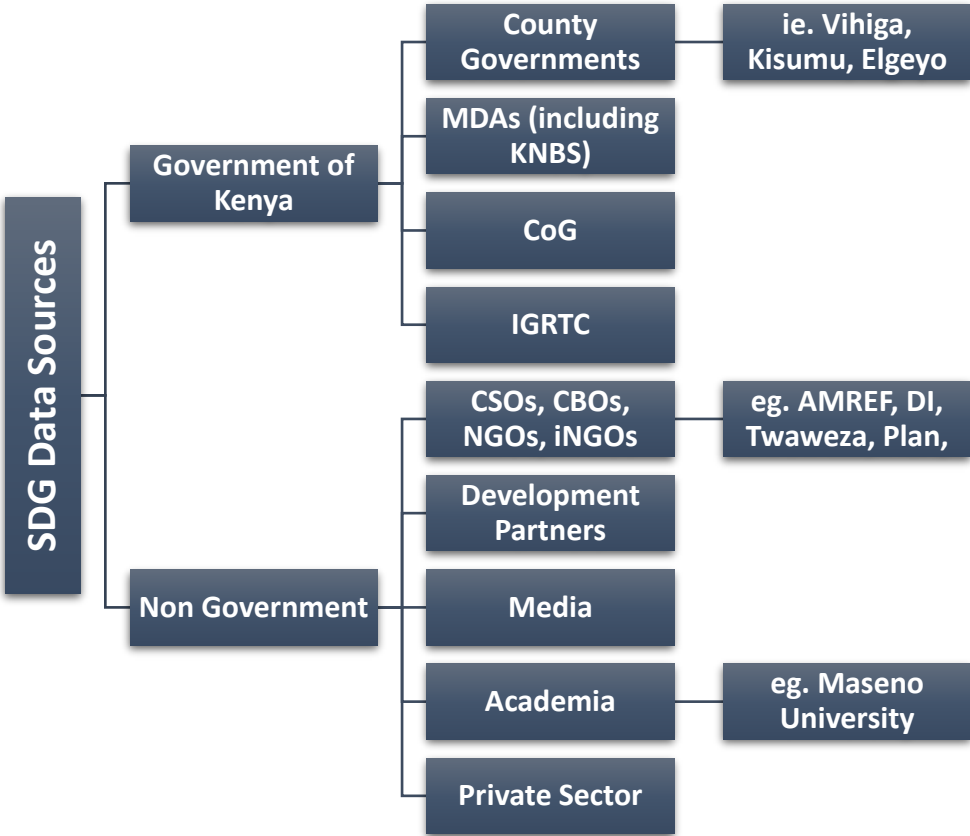


Figure 3: Current data sources for the SDGs (with some examples)

The study found that there are two broad but distinct classifications of data (based on how data is collected):

1. **Administrative data refers** to daily operational data collected by County Departments on the state of affairs, such as, roads, disease burden, enrolment. As a natural count of ALL cases of the phenomenon, this data often might not require rigorous validity and reliability tests.
2. **Survey data refers** to when a particular phenomenon is under investigation – requiring responses from individuals or households. This requires more rigor in terms of validity/reliability tests; hence sampling is a critical consideration when it comes to survey data so that the findings can be generalized to the population.

Both of the above cases are important because both kinds of data are used to influence policy and decision making for entire populations (sampling and methodology considerations are discussed in section 5.2.3).

This study did not find standardized interoperability in data across data producers. Some organizations have advanced bespoke software such as MIS that hosts M&E (outcome and impact) indicator data that is stored in relational databases. Others use Microsoft/Open Office tools, PDFs and statistical software packages - which means data can be shared with relative ease even though it will be difficult to analyse. Worse still, others keep records in manual format. In this way, there might be two organizations (or even two departments such as Water and Health) working in the same area but duplicating efforts to obtain similar data but not sharing among themselves.

The table below expounds on the sector specific data sources that inform and guide the public services provided in the respective sectors.

Table 5: Overview of SDG data sources

GOAL	DATA SOURCE
Goal 2: Zero hunger	Data in agriculture sector in Kenya is concerned with economic and other statistics related to collection, processing and analysis of data on domesticated plants and animals. It also extends to ‘rural statistics’ which is a broad range of statistics from rural areas in the country covering economic, social, demographic and agricultural domains ²⁸ . These are often coordinated and collected by the Ministry of Agriculture, Livestock and Fisheries and KNBS who also collect a lot of

²⁸ Kenya has developed a Strategic Plan for Agricultural and Rural Statistics 2015-2022. <https://www.knbs.or.ke/strategic-plan-for-agricultural-and-rural-statistics-2015-2022/>

	<p>data. These national bodies works closely with county departments (most data activities occur at the county level with the exception of agricultural research institutions) who in turn have extension officers in each subcounty and who they rely on to collect administrative data from farmers as well as through programs such as the fertilizer subsidy program</p>
<p>Goal 3: Good health and well-being</p>	<p>KNBS undertakes the Kenya Demographic & Health Survey five (5) years²⁹ (it has been indicated they do not rely on DHIS2 to avoid bias). Similar to agriculture, they also collect data from the parent Ministry (Ministry of Health). When undertaking the KIHBS, they ask about the main sources of drinking water and whether these sources are treated as well as education related information.</p>
<p>Goal 4: Quality education</p>	<p>The functions of education fall under the Ministry of Education, Science and Technology in Kenya - but they are shared between national and county governments³⁰. Thus, KNBS normally sources data from MoEST and the various county governments especially when it comes to enrolment data and the infrastructural set up. During social surveys, KNBS gathers data on the education level of respondents. It relies on performance information from Kenya National Examination Council as well as from Universities and Colleges.</p>
<p>Goal 6: Clean water and sanitation</p>	<p>Safe water coverage remains a key issue reported on by the Ministry of Water and Sanitation. The KIHBS remains an important source of information for household level information on water. Among others, they check what the main sources of drinking water are and whether the water is treated.</p>
<p>Goal 7: Affordable clean energy</p>	<p>KNBS collects data that checks proportions of households using clean energy. They investigate sources of lighting and cooking energy. In fact, in the 2015/2016 KIHBS, they even weighed firewood.</p> <p>Functions and services in this sector are not yet devolved; a lot of guidance is still provided by the Ministry of Energy - with services being provided by various institutions that fall under the State Department of Energy and State Department of Petroleum³¹ with 11 semi autonomous institutions³² providing services in this sector.</p>

²⁹ The 2014 Kenya Demographic and Health Survey (KDHS) is the most recent survey and which follows up KDHS surveys conducted in 1989, 1993, 1998, 2003, and 2008-09. KDHS 2014 was the first national survey to provide figures for demographic and health indicators at the county level (KDHS 2014).

³⁰ The functions of Educational and training are shared between the national and county governments. The functions of the National Government are: education policy, standards, curriculum, examinations, granting of university charters, universities, tertiary educational institutions, institutions of research, higher learning, primary schools, special education, secondary schools, special education institutions and promotion of sports and sports education. The functions of the County Government are: pre-primary education, village polytechnics, home-craft centres, farmers training centres and childcare facilities.

³¹ See functions http://energy.go.ke/?page_id=528

³² See a list of institutions http://energy.go.ke/?page_id=528

5.1.2 Data Collection and Validation

a. County Governments

Results: Data collection methods and validation

Vihiga: In Vihiga County, the Department of Health collects all health related data especially relating to disease occurrence and prevalence. The Department of Education collects all ECDE and TVET data, that is, the data for the devolved components of education. This includes data on the number of children enrolled in ECDE schools and TVET institutions. Other data such as transition rate from pre-school to primary and from primary school to secondary school, data on teachers in schools and their qualifications is provided by the National Government through the MoEST and KNBS. The Department of Agriculture collects data on agricultural activities in livestock, fisheries and crop development.

The Department of Environment, Water, Energy and Natural Resources in Vihiga County depends on Lake Victoria South Water Services Board for data. This is because the Board carries out water surveys in different counties around Lake Victoria.

Validation example: *The county department of health makes use of the County Health Management Team to validate the data collected through different tools placed in health institutions across the county.*

Kisumu: In Kisumu County, the Department of Health collects all health-related data. This is mostly administrative data that is collected using different data tools in the health centres. The Department of Education, ICT and Gender collects data on youth, early childhood development, transition rates for schools, cash transfers, and people living with disabilities. Because it is in their mandate, the department tries to mainstream gender not only in the activities and roles carried out in the department but across all the departments in Kisumu County.

In Kisumu, (Department of Water, Environment, Natural Resources and Climate Change), data on water facilities is collected directly by staff from primary sources. The department also liaises with Kisumu Water Services Company to get data on already existing facilities.

Validation example: *In the Department of Health and Sanitation, validation of data is done by District Health Officers. Once the data is collected, it is validated for accuracy before it is posted on the DHIS.*

Elgeyo Marakwet: In Elgeyo Marakwet County, data quality and availability depend on the respective sector. The most available as well as advanced (in terms of frequency, actual

collection, analysis, validation processes and dissemination) is Health Department Data³³; followed by the Education Department³⁴. Water³⁵ and Agriculture³⁶ data are less developed in terms of the criteria highlighted.

Elgeyo Marakwet has enacted a Public Participation Act (2014), has instituted the Office of Public Participation which facilitates and coordinates public participation in the county. This office ensures that the sessions cover the broad spectrum of public needs and interests with clear context backed by data and evidence by not only facilitating budget process but any other issue of public importance - at least once in three months.

The county department of health³⁷ makes use of the County Health Management Team to validate the data collected through different tools placed in health institutions across the county.

Validation example: *In education, Elgeyo Marakwet for example undertakes a random sampling exercise to check on the population of ECDE learners being reported. Another way used in verifying data in the Departments of Education and Agriculture (in the three counties) is through site visits to verify that what was reported is the actual situation on the ground. In agriculture, the department has validation meetings so that they validate the county level data before submission to the national level or to KNBS. Stakeholders comprise the business community (financial institutions such as banks), NGOs (such as World Vision which is active in Elgeyo Marakwet), and other CSOs representing the Youth, PWDs, women and marginalized communities.*

All counties under study: Data validation *especially when it comes to survey data is important; it mitigates against the data released being different from the actual situation on the ground, data is validated and verified before it is officially released. Departments of Health in all three counties have different data collection tools³⁸ that are placed in health institutions (including specific registers for HIV and TB). They are made use of in health facilities such as HIV care centres, outpatient centres, hospital wards and antenatal centres. And even though the data may exist in DHIS, data reviews must be done to deal with any errors after the*

³³ Disaggregated to the facility level

³⁴ The county government is able to disaggregate by age, gender and the school.

³⁵ Disaggregated only by location at the county level. The last available data in use is drawn from KIHBS 2015/16.

Whereas Kenya Environmental Sanitation and Hygiene (KESH) policy of 2016-2030 (developed by the Ministry of Health covers some of these indicators) emphasizes the systematic collection of data, this is not happening in practice on the ground.

³⁶ Disaggregated by location to sub-county level

³⁷ Data reporting for the previous month in DHIS2 is closed by the 15th of the next month so that any data issues can be discussed. In Elgeyo Marakwet, they undertake data quality audits and monthly data review meetings. The data reporting date is also a very important indicator discussed in these meetings.

³⁸ Health data is sourced from DHIS2; it is invaluable to both state and non-state stakeholders. In Elgeyo Marakwet, the SNV Netherlands Development Organisation has also established a tool that is in use for Community Led Total Sanitation (CLTS). They are also using an Electronic Medical Records (EMR) system, IQCare which has been installed by the iNGO Palladium. MTIBA system (for TB) is used to track resistance of patients – even when they move across counties.

monthly data capture has taken place.

Agriculture data validation involves county departments sitting in the same forum with the Ministry of Agriculture Livestock and Fisheries, other national level entities such as state departments, Sugar Board, Kenya Agricultural and Livestock Research Organization (KALRO), and affiliated research institutes, administrative data from county departments, and other non-state actors, who will share and assist to validate any data that has been collected by KNBS

Validating data can however be difficult especially in the education sector where pupils' population in school changes frequently (due to learners dropping out, absenteeism or change of schools).

b. Non-State Actors

At the community level, CSOs and non-state actors mostly use official data for baseline information that helps them to design programs. Implicitly data users know that this information is not current; hence users can use the data to design a program at the beginning, but the data is not useful to monitor progress during implementation. For example, if there arises a need to distribute commodities such as fertilizer in a particular location, we will need to know how many farmers there are and the sizes of their farm holding for a baseline. Once this baseline has been established, there is a need to continue monitoring and so most of the CSOs interviewed (as much as possible) undertake their own periodic collection of data to inform their programmes and activities. CSOs deploy a variety of tools (in physical or electronic format) in the field using enumerators.

Nevertheless, it was reported by KNBS that their (KNBS) involvement in the design stage for CSO data collection tools seldom occurs. The same applies to results validation after data analysis and before data has been published – by non-state actors. But this does not mean that CSOs generally do not apply rigorous validation. See the Case Study 1 from Twaweza. KNBS indicates that high-quality reports like these would get published within ambit of official statistics. Those undertaken by non-state actors have built in validity/reliability of tools at the programme and research design stage.

Case Study 1: Data validation for Uwezo Project - 2011 to 2015

(Source: Twaweza)

A good example of non-state collaboration with KNBS can be demonstrated by Twaweza in their Uwezo project (already mentioned in section 4.3). During the project design stage, the project undertakes sample development in conjunction with KNBS. They sample the Enumeration Areas (EAs) including selecting them on a map and pick 20 households randomly within each EA. Some of the definitional problems they have discussed with KNBS in developing their sample include: 'do we include households without children?' They deliberated this and eventually decided they must include households without children (so that they do not bias against such households) so that the project can withstand the test of being a truly random sample which can inform policy. Further, for the English, Mathematics and Critical Thinking tests, the project works closely with curriculum experts from Kenya Institute of Curriculum Development (KICD). During the results dissemination stage, they involved KNBS again to review the analysis of qualitative findings before publication.

Initially, the result of the study was not received well. It was viewed as an affront on literacy and numeracy in the East Africa states - particularly Uganda and Tanzania; not that Kenya was doing much better. In Uganda, the Ministry of Education referenced Uwezo findings³⁹ at a time when the teachers union was threatening to strike. But with time, these results have become relevant in raising the following policy issues⁴⁰:

1. Repeating a class or automatic transition are not a cure for lack of accountability if there are wasted resources in keeping the children in school.
2. The study opened up discussions for curriculum developers to think 'what' the children ought to be learning at their level.
3. Children have a problem grasping (reading/writing) English as the main mode of instruction.
4. Assessments should be geared towards addressing areas of difficulty which the learners should be taught; not for these to primarily inform who moves to the next class.

On the one hand, survey data should be validated; on the other hand, county government planning units argue that for administrative data (drawing from 100% of the population) validation is not necessary. The current practice is that for county data to be validated at the national level, KNBS county statistical offices report the information from the subnational to the national level for validation. At the national level, KNBS convenes Technical Working Groups (TWGs) that are sectoral in nature (encompassing Health, Education, Governance, Environment, Trade, Gender) to validate the data in question. TWGs are comprised of KNBS staff and relevant staff drawn from MDAs - with the requisite background and expertise to validate information. The main challenge remains that KNBS takes too long to validate the data, delaying county activities (like CIDP activities or public participation process).

³⁹ Ministry of Education, Uganda. Press Release on Teachers strike over 10% salary increment.

<https://www.scribd.com/document/265530060/Ministry-of-Education-Statement-on-Teachers-strike>. (viewed 10/06/2019)

⁴⁰ Twaweza (2018). *Are our children learning beyond basic skills? Lessons from Uwezo learning assessments from 2011 to 2015*. <https://twaweza.org/uploads/files/East%20Africa%20Report%202015.pdf>. (viewed 10/06/2019)

5.1.3 Barriers to Data Collection

Results: Barriers to data collection

Vihiga: Some private schools intentionally give wrong information thereby affecting planning by county government. Also, citizens fear victimization and are thus reluctant to participate in the data collection process. Political influence sometimes leads to blacklisting of some local organisations that face hostility when collecting data. Some respondents also demand for payment before they participate in data collection exercises.

Kisumu: Some respondents intentionally give wrong information thinking that all data collectors are donors. For periodic data collection, respondents become hostile and urge data collectors to use previous data collected from them. Finally, research is a very expensive exercise and inadequate resources affects the efficiency and effectiveness of the process. At times inflation of figures to suit certain agendas occurs; this is especially if the desire is political or to mobilize additional resources. This was mainly mentioned with regards to HIV/AIDS cases reporting (especially by CSOs).

All counties under study: In all the counties visited, citizens are interested in discussing data and policies that affect them at the local level, and not at ward, sub county or even county level - which is the current standard for most of the data collected. The lack of data disaggregation of data at lower levels within the counties of Vihiga, Kisumu and Elgeyo Marakwet presents challenges the enabling of public participation at the local level (especially where the priorities are public services delivery)

Below are some of the barriers cited by respondents (both county government and CSOs) to data collection:

- 1. Logistical issues** such as transport and communication. Low internet penetration slows data collection process; especially in situations where data collection tools require internet connection.
- 2. In cases where data collection is carried out frequently,** respondents urge collectors to use data collected previously for a similar study.
- 3. Respondents are do not understand that their data can be used as a political tool** (See Figure 7: Decreasing levels of trust in data from Left to Right as per interviews).
- 4. Respondents are reluctant to participate in the process without assurance on how the whole process is of benefit to them.** Sometimes respondents demand to be paid in order to be interviewed. This interferes with random sampling and will create bias. In other cases (and sometimes with merit) respondents feel that data collectors will just use their data to acquire funds from donors - which will not trickle down to the grassroots and so are reluctant to participate.

5. Lastly, respondents are also wary of negative politics - sometimes which involves inflating or deflating results - such that even if they participate, it will not truly translate to their own priorities for service delivery.

5.2 Resourcing for Data Collection

a. County Governments

Results: Funding for data collection

Vihiga: In Vihiga, the county government funds the data collection processes. Allocations for this formally occurs during the budgeting cycle. All departments have up to 2% allocation for M&E.

Kisumu: The county government budget is the main source of funds for data collection processes. However, donors and other partners (CSOs and academia) carry out data collection activities that supplement data from the county government.

Elgeyo Marakwet: Mostly, data is collected through M&E functions within departments, as funded by the county government, to monitor the progress of implementation of the CIDP (and by extension the SDGs).

Even though county government departments draw resources through the county budget (integrated into the ADPs and CIDPs), these resources are sometimes not enough⁴¹. When it comes to data collection in the three counties, there may not be explicit IFMIS Classification codes used by departments for budgeting that will be referenced as “Data Collection”. What is known is that counties have monitoring structures that are funded and staffed through county budgets to ensure that targets are being met on project implementation and service delivery.

Nevertheless, human resource remains a challenge for effective data collection, processing, analysis and dissemination. Some counties grapple harder with this issue than others. For instance, Vihiga had 6 economists when devolution started; in contrast, Kisumu did not have any. Still, all the counties felt that more resources ought to be allocated to hire enumerators to ensure better and regular data collection. Also, more ought to be done as far as acquiring

⁴¹ Other forms of resources normally provided to the county governments are in skilled personnel who can be seconded by their respective organizations to support data activities. Other support often provided comes in the form of laptops to better the outputs of collection, analysis and curation of data.

the ‘tools of trade’ (laptops, internet connectivity, vehicles or motorbikes to facilitate easy travel to remote locations for data collection) so that when skills have been acquired, there is continuous supply of quality data. Also, county governments often rely on National Government Statisticians for data analysis; sometimes they hire short-term statisticians from other organisations⁴².

b. Non-State Actors

Many CSOs (especially at the county level) grapple with the issue of resourcing for data collection because of lack of funds. Grants are not always sustained or renewed in subsequent years which makes outcome indicator level results difficult to track continuously. In this regard, CSOs such as Twaweza do not undertake project based work; rather they have developed a strategy⁴³ (currently the 2019-2022 strategy). In recognizing the challenges of resourcing in health, for their 2018-2022 strategy, AMREF has 3 main pillars: 1) Human resource for health (capacity development) and management, 2) innovative service delivery in health (increase service access through technology across the entire spectrum of their services) and 3) improving healthcare financing (through increased county budget allocations to health). From these 3 focus areas, human resource and financing are prioritized and using public private partnerships is one such example; but it is out of the scope of this study to delve deeper into this topic.

5.3 Data Collaboration

KNBS remains an important player in the data ecosystem and both county governments and CSOs collaborate with this agency. Many institutions, especially County Government Departments, collaborate with KNBS county statisticians because of their expertise in methodology, data collection and analysis. Data validation issues have already been discussed in section 5.1.2. In all this, KNBS must remain non-partisan and committed to ensuring that correct facts and proper analysis are disseminated to the citizens.

a. County Governments

⁴² A similar example from the national level is reported by KNBS who since publishing the County Statistical Abstract in 2015, have not done any other similar publication. The plan back then was to undertake the county abstracts annually - but this has not happened because of limited resources as well as limited skilled staff.

⁴³ <https://twaweza.org/go/twaweza-strategy-2019-2022>

Results: Funding for data collection

Vihiga: County departments mostly collaborate with KNBS in terms of ensuring standards during sampling, data collection and analysis as well as during data validation.

Kisumu: For example, the Department of Education, Gender and ICT also collaborates with CoG and UN Women as these two institutions also share data with the department. Finally, there is collaboration with local CSOs. This is an important collaboration because these CSOs play a crucial role of disseminating and educating citizens on different aspects of development.

Elgeyo Marakwet: Like Vihiga, in Elgeyo Marakwet, most of the collaboration occurs with KNBS to ensure that the data meets its standards. The main issue cited by Elgeyo Marakwet when it comes to collaboration, there is a need for improved capacity especially among the non-state actors. Media for example, need to learn how to delve deeper into other data sources in order to check their stories against data and evidence before publication.

All counties under study: Overall, KNBS is the key government agency that different county departments collaborate with due to its expertise in conducting surveys that are important to supplement data collected by the county governments and ensuring that methodological rigour can support the findings from the data.

One product of collaboration between county governments and KNBS is the County Statistical Abstracts - in which KNBS is publishing data with the respective county government departments. This includes administrative data (county commissioners also play a coordinating role for departments which are represented at the national level) and other information compiled by Ministries, Departments and Agencies (MDAs) as well as KNBS Census data. The very first County Statistical Abstracts were published in 2015 and intended to be a single source of indicators covering various sectors of the economy at the county level. An important result from the process of compiling the County Statistical Abstracts is that:

1. There are guidelines that standardize data across counties for data comparability.
2. There has been emphasis on disaggregated statistics at both devolved and national levels.
3. The issue of timeliness, availability and reliability of data produced within the county itself has been put in focus so that the quality of future statistical abstracts is ensured.

KNBS also interacts with the CoG⁴⁴ and with the IGRTC⁴⁵. As an example, for the upcoming 2019 census, CoG is a member of the national steering committee of the census. KNBS has

⁴⁴ CoG draws its mandate from Section 19 of the Intergovernmental Relations Act (IGRA 2012)

⁴⁵ IGRTC draws its mandate from a) The Constitution of Kenya 2010 articles.1, 6, 174, 186, 187, & 189 b) Legislation; CG Act, 2012, IGR Act, 2012 & PFM Act, 2012, UA&C. Act, 2012

also established county census committees⁴⁶ which comprise of County Commissioners, the County Secretary as well as non-state actors in the communities like churches and CBOs.

b. Non-State Actors

The most important technical point of convergence for non-state actors work with KNBS is rigorous methodology; presently work undertaken by CSOs that can be subjected to successful tests of validity and reliability can be published officially (see Twaweza’s Uwezo case study). Importantly, KNBS is working on guidelines that inform non-state actors on methodology.

5.4 Data Publication

Results: Data publication by County Governments and CSOs

Vihiga: In Vihiga County, the Department of Environment, Water, Energy and Natural Resources, water data is uploaded to Water Regulation Information System (WARIS⁴⁷). WARIS is a database application Software established to facilitate data entry, aggregation and evaluation of data. The data is entered by Water Service Providers (WSPs). The system then generates a report based on data entered which is accessible to both WSPs and Water Services Regulatory Board (WASREB). The data collected through the system is validated with inspection reports, tariff information and annual licensee reports before it is published. WASREB⁴⁸ is a body established under the Water Act of 2016 with the main objective to protect the interests and rights of consumers in the provision of water services, while ensuring other stakeholders’ interests are also safeguarded. WASREB hosts WARIS and uses it to monitor progress and implementation of water strategies by different WSPs and make regulations on water services.

All counties under study: for all 3 counties, the Health Department publishes its data in DHIS 2. All other sectoral reports are uploaded onto the county website or stored in county government offices in hard-copy for access on-need basis.

Use of websites: CSOs, KNBS, county departments, media and academia upload data collected to their websites from which users are able to access the information. The data is mainly presented in form of reports that are published as PDFs on quarterly or yearly basis. The Kenya Open Data Initiative (www.opendata.go.ke) is also another location where data is published.

⁴⁶ County census committees will be in charge of security issues. For example, they are addressing the issue of a surge in fraudsters offering fake job placement in census activities <https://www.knbs.or.ke/census-frequently-asked-questions-faqs/>.

⁴⁷ <https://waris.wasreb.go.ke/Home/Login?ReturnUrl=%2f>

⁴⁸ WASREB sets standards and enforces regulations that guide the water sector in not only ensuring that consumers are protected and have access to efficient, affordable and sustainable services. It also facilitates the financial sustainability of Water Service Providers.

By law, all KNBS reports are published in hard and soft copy on the KNBS website⁴⁹

Printed copies - this happens especially during the budget process where some copies are availed to citizens to facilitate deliberation. Similarly, any user interested in specific data may physically visit the offices or libraries where the data is stored.

Note that one of the big challenges in data dissemination is the inability to reach multiple populations (rural, urban) using the right kinds of media because of limited financial resources.

Some data publication is very sensitive; as an example, when it comes to food security, there are protocols that only the Office of the President (due to the Kenyan historical political context) officially makes announcements. Equally when there are contagious diseases (e.g. cholera or measles), only KEMRI can confirm an incident because they are responsible for testing and following protocols around this information. This is not just to avoid causing panic, but also because there is some statistics that by law, can only be published by the mandated MDAs. Also, publishing of individual data may breach confidentiality - when it comes to research topics such as HIV, research involving questioning children. Nevertheless, most of the data being collected is made available for publication or on request.

The figure below outlines KNBS Service Charter - it is a list stipulating the timelines for the publication of different surveys.

KNBS SERVICE CHARTER		
SERVICES AND PRODUCTS	TIMELINES	CHARGES
1. Consumer Price Index	Last working day of the Month	Free
2. Leading economic indicators	Monthly	Free
3. Quarterly GDP Report	April 30 th , June 30 th , Sept 30 th , Dec 31 st	Free
4. Annual Economic Survey	April 30 th	At cost
5. Kenya Facts and Figures	May 15 th	At cost
6. Statistical Abstract	September 30 th	At cost
7. Population Census Data	Once every 10 years	At cost
8. Cartographic maps	On request	At cost

⁴⁹ <https://www.knbs.or.ke>

9. Technical advise on official statistics	On request	At cost
10. Statistical research services	On request	At cost
11. Basic statistical data	Available on www.knbs.or.ke	Free
12. Library	Monday to Friday 9am to 4.30pm	At cost

Figure 4: KNBS service charter⁵⁰

5.5 Data Skills Gaps Across State and Non-State Actors

Results: Data skills gaps

Vihiga: Data collection, analysis, dissemination and storage require skilled personnel to undertake all the activities. Lack of enough skilled personnel leads to question on validity of data that is published. Different departments face the challenge of lack of enough skilled personnel to undertake data collection. It is therefore crucial to invest in training of data collectors and analysts to enhance the validity of data published. The main challenge has been lack of adequate resources to employ enough staff as well as capacity building and equipping of these staff with skills and tools to manage different cycles of data processing. Department of water faces a problem of staff shortages, especially upon retirement of staff members. Lack of skilled and trained data collectors can lead to wrong interpretation of questionnaires and other tools used in data collection thus distorting the output from data collection and analysis exercise.

Kisumu: There is a shortage of statisticians affects working of Finance and Economic Department. The county has to rely on external statisticians mainly from the national government to assist especially in health analysis. As is the case with Vihiga County, there is a shortage of skilled personnel. In-house statisticians would be best placed to help in streamlining the data ecosystem at the county level especially analysis in a sustainable manner.

Elgeyo Marakwet: In Agriculture, due to lack of expertise, it was expressed that real market research was not happening and which may have dire implications on accurate the food situation if the department is relying on “guestimates”. For example on trends for maize based on area under cultivation, the data may state that there is enough maize at the beginning of the year but the National Cereals and Product Board (NCPB) do not have enough implied reserves.

⁵⁰ <https://www.knbs.or.ke/download/service-delivery-charter/>

All counties under study: Across departments interviewed, when they have gone ahead to train staff, the lack of resources (manifested as lack of ‘tools of trade’) impedes the continuous flow of quality data.

Despite efforts to ensure as much SDG implementation data is collected as possible, there are still data gaps owing to lack of skills to collect information. See some examples on data gaps owing to inability (in both state and non-state actors) in operationalizing some of the very important SDG Indicators (that is only proxy indicators can be used) covered in this study:

1. Due to lack of know-how, KNBS indicates they are only monitoring 4 (out of 6) indicators on clean energy sources (Goal 7)
2. Regular collection of data on the number of children who have attained school-age but are not in school is conducted; but only Twaweza is doing this (and that on an annual basis)
3. Regular update of information on birth rates is also difficult - especially because not all births happen in health facilities
4. KNBS is not conducting disaggregation of data based on gender, people with disabilities, very localized location information of citizens, counting homeless families living in the streets, for common data sets such as water and health sanitation and agricultural production

Sometimes, those involved in the data collection, analysis and dissemination (at the county government level) are not trained to do so. In all 3 counties studied, the water sector is especially affected by lack of skills (at department level) and the capacity to analyse data. Consequently, lack of in-house statisticians at the county level has led to counties hiring consultants to help analysing data (like in the case of the Health Department in Kisumu county).

Lastly, data visualization skills, hence the ability to simplify the analysis and present it in charts, graphs and maps that can be easily understood is also missing. Even CSOs like Twaweza believe they have collected a lot of data - published on their website - but it remains underutilized due to lack of visualization.

5.6 New Ways of Data Collection and Management

Traditional data collection employs paper/based data collection - whether through interviews, FGDs, among others. Technology is an enabler that can bridge data gaps in terms of data timeliness, quantity and quality.

This study finds that ICT (mobile phone technology, the use of ICT including big data, and the use of ICT for Citizen Generated Data) has not been leveraged enough - but this is basically before the 'old' methods seemingly still work. There are just 'pockets' of success that can be reported. Whereas the potential for ICT is acknowledged, three main kinds of challenges are recorded during interviews.

Firstly, infrastructure related challenges have to do with the physical hardware and software configuration that facilitates all states of data collection and processing. Unfortunately, the costs related to ICT infrastructure are prohibitive to set up and maintain. From the government perspective, the national government through the ICT Authority has the National Optic Fibre Backbone (NOFBI) project which aims at ensuring connectivity in all the 47 counties of Kenya would also be vital in easing communication across counties as well as improve access to information as well as government service delivery. County governments are also making investments in internet connectivity at the ward level through information centres or 'kiosks'. At the time of writing, less than half of all the wards in Vihiga, Kisumu and Elgeyo Marakwet are covered.

From the non-state perspective, CSOs, NGOs and iNGOs as well as academia's model entails developing a proof of concept collaboratively with the government with the view that ownership can be sustainably transferred to the government at some point - but this is not always guaranteed. Without finance, activities cannot be scaled up or institutionalized because non-state actors often do not have the funds to invest in ICT infrastructure (since ICT is an enabler, not an end in itself) beyond a ward or county.

Secondly, stakeholders are facing **interoperability** challenges related to putting in place data standards and ensuring that all actors operating in the same area are collecting similar data types which are easy to aggregate/disaggregate and to analyze. As stated in section 5.1.1 - data types are disjointed causing practical problems in bringing data together that covers the same population but has been collected by different stakeholders.

Thirdly is **data illiteracy**; this is not so much the lack of access to ICT ("material access" barrier) but the lack of proper knowledge, education on ICT that ensures that no one is left behind. With ICT, there is a real risk that the divide widens between the "information rich" and the "information poor" which would exacerbate the problems of social and economic injustice. Successful deployment of ICT by state and non-state actors must create new opportunities for those who are traditionally isolated from mainstream programming; in particular noting, that ICT programs often leave behind women, children, persons-with-disability, immigrants and the elderly.

5.6.1 Use of Mobile Technologies

a. County Governments

The use of mobile technologies by county governments has mostly taken the form of citizen reporting via phone calls.

Results: Use of mobile phones for data collection

Vihiga: Vihiga County has 0800720378 as its toll free line - and which is used as a data collection tool. This line is used for phone calls where citizens can report among other issues suspected disease outbreaks, gender based violence, insecurity issues, whistleblowing on corruption and any other issue that citizens deem urgent to be addressed by the government.

Kisumu: Mobile phones are used mainly by citizens to report on any arising issues that needs government intervention. This is in situations such as Gender Based Violence (GBV), suspected disease outbreak, security issues and reporting on lack or inadequacy of social services such as water.

Elgeyo Marakwet: In Elgeyo Marakwet, there is a desk stationed at Huduma Centre in Iten where citizens can report issues. Presently, this channel is mostly used by tenderers seeking clarifications or getting the status of their payment process (for example, responses to procurement questions is done once a week on Wednesdays).

At the national level, there are innovations at the Ministry of Health which have been piloted to unlock bottlenecks in data flowing from the community health level as per the case study below. This implementation was undertaken in Siaya and Homabay counties. It is likely going to be scaled in Kisumu county as well given the positive results so far.

Case study 2: UNICEF Program Monitoring Response (PMR) - Kenya (Source: UNICEF East and Southern Africa Regional Office⁵¹)

With (PMR just being established after county governments were formed) respect to administrative structures in Health, Kenya considers the lowest level (tier 1) of service delivery in the health sector

⁵¹ UNICEF ESARO (2017). Synthesis Report: Evaluation of ESAR Institutional Strengthening Support Initiative on Decentralized Programme Monitoring and Response

https://www.unicef.org/evaldatabase/files/ESAR_PMR_Evaluation_Synthesis_Report_FINAL_29_JUL_17.pdf (viewed 04/06/2019)

In 2014, UNICEF headquarters, in collaboration with its East and Southern Africa Regional Office (ESARO), sought to implement a regional institutional strengthening support initiative to improve the Programme Monitoring and Response (PMR) capacity of the centralized government decision-making bodies and communities in four countries (Kenya, Swaziland, Uganda and Zimbabwe). They would do so by addressing targeted issues related to the health of women and children. Through the programme, UNICEF Kenya Country Office has had an overall focus on Health, Nutrition, WASH and HIV. It has provided targeted technical assistance and support to a number of communities in Siaya, Turkana, Garissa, Homa Bay, Kakamega, and Nairobi.

to be the household level, followed by the community (tier 2), facility (tier 3), sub-county (tier 4) and the county (tier 5). For PMR, UNICEF KCO partnered with the University of Nairobi - School of Computing and Informatics, MOH with participation of Homa bay and Siaya Counties' Departments of Health to extend functionality of DHIS2 to accommodate household level data collection.

Specifically, health at community level was targeted with the aim to identify and promote good practices by health managers and community service providers at sub-national levels to make data available on a near real-time basis for result indicators related to the delivery of essential services (health, nutrition, HIV, etc.). PMR in Kenya aimed to:

1. Make data available on a near real-time basis for select intermediate result indicators related to the delivery of essential services;
2. Strengthen links between citizen engagement, social accountability and service delivery;
3. Enhance the use of data by sub-national decision-makers to facilitate action aimed at reducing barriers and bottlenecks to service delivery;
4. Establish and/or refine knowledge exchange mechanisms for in-country and cross-country learning;
5. Document and evaluate demonstrated approaches to decentralized program monitoring and response to inform national scale-up of successful models.

In Kenya, the PMR project did not really change the established data collection and aggregation systems in the health sector. The unique contribution of the PMR initiative was enhancement of county HMISs. There are several stakeholders and processes/systems/infrastructures that existed at the county level. ICHIS was an enhancement of DHIS2 at the county level that made sure that there is a coherent process to make data useful at the national level. This is outlined in the diagram below.

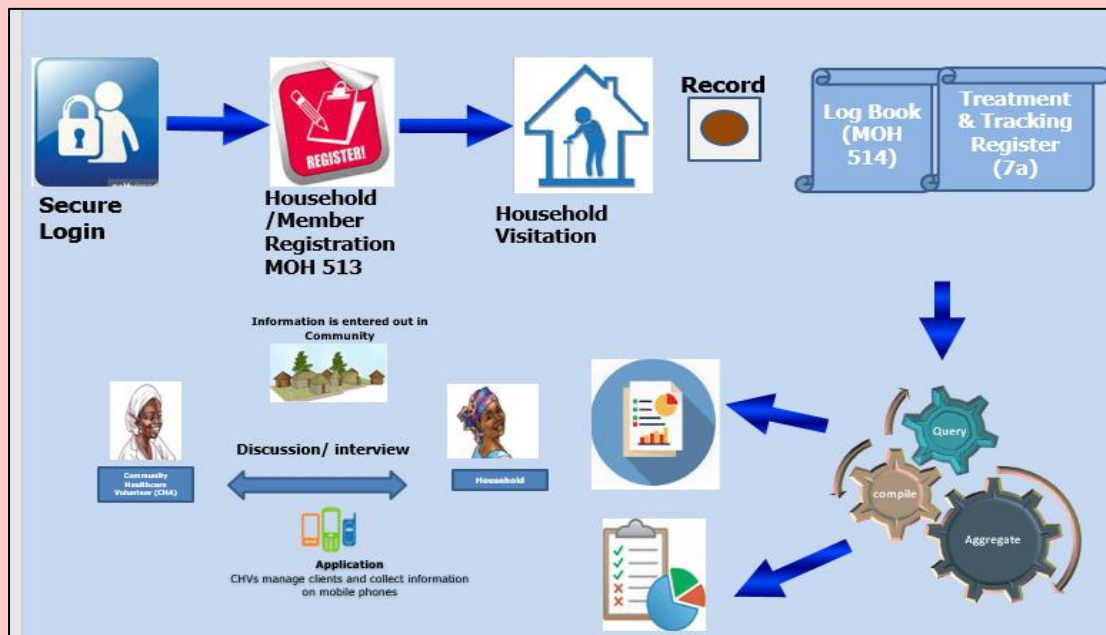


Figure 5: ICHIS Process/Data Flow.

Electronic records are generated to assess community health units (the MOH 513⁵² is used for household registration, MOH 514⁵³, MOH 515⁵⁴ and MOH 516⁵⁵ and the RMNCH scorecard⁵⁶) functionality; and these are aggregated by ICHIS (an enhanced DHIS2 instance at the county level) before transmission to the national level DHIS2, where they can be accessed in real-time. More importantly, the granular data in ICHIS is available for access via a mobile phone or tablet through an internet connection and it is used to facilitate community dialogues and other feedback loops that improve accountability in service delivery.

COMMUNITY HEALTH UNIT CHALK BOARD														
YEAR: 2017				ORG UNIT NAME: Homa Bay County										
Indicator	No.	Indicator	No.											
Total population	5881.0	Total adolescent and youth: Girls (13 - 24 years)	312.0											
Total households	1928.0	Total adolescent and youth: Boys (13 - 24 years)	350.0											
Total villages		Total elderly (60+ years)	162											
Total women 15-49 years	897.0	No. of household with functional latrines	203.0											
Total pregnant women	3.0	No. of households with hand washing facilities	116.0											
Total children 0-6 months		No. of children fully immunized												
Total children 0-11 months old		No. of household using LLINs												
Total children 0-59 months old														
Indicators	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total	
No. of households visited		1928.0											1928.0	
No. of households using treated water		941.0											941.0	

Figure 6: Community Health summary data for MoH 516 from ICHIS

Before the PMR initiative, the data collection was bottom-up; there was no real responsibility around the quality of what was recorded because the CHVs largely perceived their role as data producers; they did not understand the importance of indicators and how data is used. This was also implicitly difficult to analyze because data resided in the manual registers. The PMR initiative has changed this; by availing data on mobile, it is now possible to have community health dialogue days focused on discussions based on data and evidence. This is also because PMR has built capacities to analyse and interpret the data at the tier 1 level in the pilot counties - with the net result effect of stimulating the use of these tools and stimulating demand for data.

b. Non-State Actors

CSOs interviewed often use a mix - some of the surveys are done using mobile phones but the bulk of them are still undertaken manually. Tools range from open source to closed tools;

⁵² MOH 513: The household register is a record where we write major household events or services at the household registration and after every six months. The Head of the Household should be able to respond and give detailed information about basic information on what was done or identified in the household to measure the actual CHW's outputs and outcomes as a result of household visitations.

⁵³ MOH 514: The Service Delivery Log Book is a diary that is used to collect information on what was done or identified in the community, among households and/or individuals served from the household during the period of offering a health service, health messages or defaulter traced. The Service Delivery Log Book gives the numerator for measuring the effort of the caregiver for the previous month.

⁵⁴ MOH 515: This tool is the monthly summary of the CHWs efforts and services offered at the household level. The information collected measures the CHW's efforts and services offered at the household level. It shows the Community Health Unit (CHU) Outputs.

⁵⁵ MOH 516: This the current community health unit chalkboard.

⁵⁶ The scorecard is a broad reaching but dynamic tool that summarizes services across services (HIV, TB). However, it can be customized further

some of the most common apps in use to collect such data include SMS⁵⁷, USSD⁵⁸, ODK⁵⁹, ONA⁶⁰, among others.

5.6.2 Use of Centralized Management Information Systems (MIS)

a. County Governments

In some cases, for example in the health data collection, citizens are ‘expected’ to give their information including geographical information, contact information and family history (that is, number of children) in order to be provided services. This information is stored and analysed to help in planning and implementation of interventions.

Results: Management Information Systems in use at the county

Vihiga: Vihiga County has launched a GIS data platform that will act as a central data reference platform for all data for the County. This platform will ensure that users are able to access and consume data of different sectors in Vihiga County. This platform will also inform the government of where to put up facilities such as hospitals and ECDs using satellite data.

For the water sector, WARIS is the centralised MIS in the water sector that is managed by WASREB – and that was found to be in use in Vihiga county for the named reasons.

Kisumu: The county is working on creating a Central Data Operations Unit. This unit will provide a platform where all the county data will be keyed in and can be accessed. This will provide a central platform that stakeholders will use to compare data over certain periods and also for budgetary planning by county departments. Also, in Kisumu (though not centralized) a water management system exists. Other sectors, such as water and sanitation, energy - do not yet have a centralized MIS⁶¹

⁵⁷ Short Message Service (SMS) is a text messaging service component of most telephone, internet, and mobile-device system that uses communication protocols to enable mobile devices to exchange short text messages.

⁵⁸ Unstructured Supplementary Service Data (USSD) is a system for mobile communication technology that is used to send text between a mobile phone and an application program in the network. It allows chat-like behaviour but only for a limited time called a session.

⁵⁹ The Open Data Kit (ODK) is a free and open-source software for collecting, managing, and using data in resource-constrained environments for it allows the collection of data offline and submit the data, when internet connectivity is available

⁶⁰ Ona provides a web and mobile app that allows the monitoring of real time field data. It enables mobile data collection in real-time. <https://ona.io/home/>

⁶¹ Centralizing and automating functions almost always leads to: 1) increased government efficiency, 2) reduced citizen bureaucracy (no need to visit multiple offices); 3) empowered citizens especially by making once opaque processes that depended on the officer offering the service now transparent 4) Through automation of services (eg. e-procurement) the private sector can be influenced to modernize as well.

All the counties under study (including Elgeyo Marakwet): In Elgeyo, Kisumu and Vihiga Counties, the Departments of Health upload their data to the DHIS2. Thus, the health sector is covered by a centralized MIS. In agriculture, there are tools like the Kenya National Livestock Management Information System⁶², Plant and Animal Health Factsheets which has simple pictorial information for quick reference on plant and animal health factsheets⁶³, livestock breeds and crop varieties suitable for various agro-ecological zones in the country, amongst others. However, these have not scaled up for national use.

One of the aims of WASREB is to bring efficient services closer to all the people (especially the poor) whilst still holding the service providers to account. This cannot be achieved without monitoring key water indicators (Water Coverage, Drinking Water Quality, Hours of Supply, Non- Revenue Water, Metering Ratio, Staff Productivity, Staff per 1000 Connections, Personnel expenditure as a percentage of Operation and Maintenance Costs, Revenue Collection Efficiency, Operation and Maintenance Cost Coverage, Sewerage).

The MoEST has also deployed NEMIS to be the source of quality, reliable and timely education data to enrich planning for the Kenyan Education System.

b. Non-State Actors

With the exception of AMREF (also working in the health sector) which has advanced centralized MIS for M&E data curation and analytics for indicators at outcome and impact level, CSOs lack centralized MIS to store and manage monitoring data.

5.6.3 Use of Citizen Generated Data

a. County Governments

Results: Citizen Generated Data

Vihiga: Citizen Generated Data in Vihiga County takes the form of data collected through the County's toll free service where citizens call to report on incidents they deem urgent for address by the county. Furthermore, CGD is also important in collecting data on Gender Based Violence (GBV). Using the toll-free number, citizens can provide data in case of complaints or issues that they want to report.

Kisumu: Kisumu County has *rumor logs*. These are books that store information coming from citizens. This information is however not taken as the real occurrence but just as a rumour

⁶² <http://www.lmiske.go.ke/lmis/home.htm?action=getData>

⁶³ <http://www.kalro.org/alris/index.php/home/emimilistall>

until the matter has been investigated. An example given was the suspicion of a cholera outbreak where citizens report on suspected cholera outbreaks (also see section 5.1.5 on data publication)⁶⁴.

Elgeyo Marakwet: As stated, in Elgeyo Marakwet, there is a help desk stationed at Huduma Centre in Iten, Elgeyo Marakwet, for this.

Citizen Generated Data⁶⁵ (CGD) has been a common practice in the three counties. This is mainly in cases where citizens are reporting emergencies or complains.

b. Non-State Actors

The work in CGD originates from CSOs as data producers. A good example of CGD at scale is Twaweza's annual large-scale report which is collecting data from random sources using citizen volunteers⁶⁶.

Case Study 3: Citizen Generated Data at scale. Uwezo Project influence in 2013 (Source: Twaweza)

Uwezo is an initiative seeking to improve literacy and numeracy in Kenya, Uganda and Tanzania. Its focus is to ask critical questions on the quality of education, and through citizen generated data that empowers parents, students and local communities advocate for policy changes that will lead to improved education outcomes.

Uwezo was coordinated by the help of more than 20,000 citizen volunteers in 362 districts across East Africa, collecting data from more than 100,000 households and 274,335 children annually. The main finding (in 2013/14) was that most children were not acquiring basic reading skills despite going to school.

One of the primary outcomes of the study was the shift in education policy discussions remaining the preserve of few academics and policy makers at the ministries of education; the space for discussion and criticism on education was widened. There has been stimulation on public policy debates specifically on the quality of education.

⁶⁴ See example of KEMRI 11TH Annual neglected tropical diseases conference (2017) citing data from the rumor logs

⁶⁵ Citizen-generated data is data that people or their organisations produce to directly monitor, demand or drive change on issues that affect them. It is actively given by citizens, providing direct representations of their perspectives and an alternative to datasets collected by governments or international institutions. (Civicus,2017)

⁶⁶ <https://twaweza.org/uploads/files/Monitoringbrief3-Uwezo140514ty+rc+vl.pdf>

Case Study 4: Global Fund Malaria Project (Source: AMREF Health Africa)

Through the Global Fund grant, AMREF has been implementing community case management of malaria in western Kenya since 2012. The counties covered include Bungoma, Busia, Kakamega, Vihiga, Kisumu, Siaya, Homabay, Migori as well as Kisii and Nyamira. Community case management of malaria (CCMM) is one of the new approaches adopted by malaria endemic countries to reduce the burden of malaria for vulnerable populations. The main premise treats CHVs as first responders, who when well trained on CCMM are adequately able to respond to fever within 24 hours- reducing mortality and morbidity. The overall objective is to contribute to the national goal of reducing malaria incidence and deaths - in line with the Kenya Malaria Strategic Plan 2019-2023. Some of the key interventions include:

- 1.** Community case management of malaria (CCMM) - key activities include training and support supervision to CHVs who then test and treat uncomplicated malaria at community level.
- 2.** Health Management Information System (HMIS) - key activities are data quality audit (DQA) and health facility supportive supervision (HFSS). CHVs and CHEWs are involved in data collection activities.
- 3.** Health Systems Strengthening (HSS) - key activities include Community Health Units formation and incentive payment to CHVs

The CHVs through the CHEWs get malaria commodities from link facilities for use in CCMM and account for the commodities by submission of community monthly reports to the CHEWs who eventually summarize the reports and share them with facility in-charges for onward submission to sub county health records and information officers (SCHRIO) for uploading onto the District Health Information System (DHIS).

The investment in community case management of malaria (health seeking behavior and now consider CHVs as their first point of contact whenever they are sick) has increased community members' access to malaria control interventions. This has probably contributed to the reduction in the prevalence rate of malaria in the lake endemic zone. In the 2015 Kenya Malaria Indicator Survey (KMIS 2015), the malaria prevalence in the lake endemic zone was 27% having dropped from 38% in 2010.

5.6.4 Use of Big Data

The study reveals that whilst appreciation of the potential for big data⁶⁷ exists in theory, it is not well understood. Respondents (both state and non-state actors) had very little or no information on big data.

⁶⁷ Big Data (such as satellite imagery or call detail records) in this study is described as data sets that are impossible to store and process using common software tools, regardless of the computing power or the physical storage at hand. Big data can provide snapshots of the well-being of populations with high levels of granularity thereby helping to discover anomalous changes that may serve as proxies for underlying well-being.

An exception is Vihiga County, which in practice, has set up a GIS data lab that will act as a central data reference platform for all data for the County. With support from France-based Airbus Defence and Space⁶⁸, US-based ESRI⁶⁹ and Kenya-based LocateIT⁷⁰, Vihiga has deployed space technologies and geographic information systems (GIS) to facilitate planning and management. Launched on 16th May 2019, the lab will assist the county government distribute resources evenly, coordinate and improve communication, visibility, accurate revenue collection and mitigate historical land injustices.

5.7 Anonymization of Data

Results: Ensuring Anonymity and confidentiality of citizen's data

All counties under study (Vihiga, Kisumu and Elgeyo Marakwet): Data anonymization is guaranteed because different departments release general county figures mainly in number or percentages as opposed to individuals' data. However, due to suspicion and reluctance by citizens to participate in the data collection due to lack of trust that their personal information will be anonymous, data collectors have to seek consent from citizens. Secondly, they have to assure citizens that personal information will be anonymized. Finally, citizens have to be assured that the data collected is used to bring about positive outcomes for those interviewed.

Consent is sought before data is collected from citizens. However, there are exceptions for data collection tools present in health facilities. Whereas personal information such as name, age, location among others are collected, when it comes to sharing of data, it must be anonymised and aggregated in the best way that avoids identifying of individuals.

Both state and non-state actors seek consent from citizens during data collection. Academia and CSOs are also required to obtain a research permit⁷¹ from the National Council of Science and Technology (see application guidelines and costs from NACOSTI⁷²). Ethical considerations must be adhered to by all researchers; the most basic of this is informed consent, especially when it comes to interviewing children or examining individual patient records. For example, KEMRI has a Scientific and Ethics Review Unit whose sole purpose is to effectively and robustly facilitate research with human participants and to protect their rights and welfare⁷³. Similarly,

⁶⁸ <https://www.airbus.com/space.html>

⁶⁹ <https://www.esri.com/en-us/home>

⁷⁰ <https://locateit.co.ke/>

⁷¹ Every research project that is carried out in the country requires a permit approval. This also includes academic research carried out by master's thesis, a doctoral dissertation, or independent faculty researchers.

⁷² <https://research-portal.nacosti.go.ke/ApplicationGuidelines.html>

⁷³ <https://www.kemri.org/index.php/blog/2015-07-06-18-01-18>

non-government institutions like African Population and Health Research Centre (APHRC) also have a policy document that outlines the responsibilities of staff who are not only generating but also disseminating research evidence⁷⁴. Sometimes citizens are reluctant to give information. Citizens' reluctance to fill questionnaires or to be interviewed is based on fear of victimization in the event they provide sensitive information. Also, citizens often demand to know what the data will be used for and how they will benefit from the whole data collection exercise.

5.8 Politics of Data

Politics also play a key role in the data ecosystem given that when a county government is formed, the government has a manifesto to implement (which may not necessarily have been developed based on data-based priorities). On the other hand, KNBS must drive a higher agenda, one that is committed to facts and that emphasizes all sectors (not just the ones the counties might be focused on). In fact, in developing the county abstracts, standardizing information was a major achievement. In 2015, they sent national coordinators who went to the counties to work with the county governments to collect data; most of the work was done in the counties with compilation at the national level. The county abstracts became a useful tool that has informed planning (including development of the CIDP).

Case Study 5: What happened to these Mega Projects in Kisumu? (Source: Daily Nation⁷⁵, Kisumu County Integrated Development Plan 2013-2017⁷⁶)

In the first generation CIDP for Kisumu, water transport improvement was one of the proposed projects under the road transport sub sector. Activities included dredging of the mouth of River Nyando to open up the river for river transport⁷⁷; introduction of ferries and ships and managing the Hyacinth. In 2013, Kisumu County, had planned Sh400 billion ship that was to be built by a Canadian company. The company signed a pact with the county government in October 2013, to build 22 ferries that would ply the various ports on Lake Victoria; but there has been no progress on this at the time of writing more than 5 years later (Note that recently, a ground-breaking ceremony was done in February 2019 by H.E. Governor Nyong'o to commence construction of a ship yard by a Chinese maritime company⁷⁸).

⁷⁴<https://aphrc.org/wp-content/uploads/2019/02/GUIDELINES-ON-PROPOSAL-DEVELOPMENT-DATA-USE-PUBLICATIONS-RESEARCH-ETHICS-AND-ETHICAL-REVIEW.pdf>

⁷⁵Odhiambo, M., Chepkoech, A. (2014). 'Failed' mega projects put 4 governors on the spot. Daily Nation. <https://www.nation.co.ke/counties/kisumu/Failed-mega-projects-put-4-governors-on-the-spot/1954182-2633184-cnxrqmz/index.html> (viewed 4/06/2019)

⁷⁶ <https://cog.go.ke/downloads/category/82-county-integrated-development-plans-2013-2017>

⁷⁷ East African (February, 2019) It was reported in January 2019 that a dredging machine had docked and work was underway to dredge Kisumu port to ease navigation of vessels; an exercise that would take 30 months. (viewed 4/06/2019)

⁷⁸ <https://www.standardmedia.co.ke/article/2001312972/nyong-o-to-launch-ship-assembly-yard>

These are examples of when poor or no data takes precedence in decision making. It also demonstrates some of the problems that may arise when data has not only demonstrated viability, but also feasibility of public projects. Will citizens utilize these public projects, especially when they have not been involved or consulted?

Politicization can influence data collection to serve special interests (whether of state or non-state actors). Consequently, government departments tend to trust only data that comes from KNBS and other MDAs. The government does not trust the media. Media is seen as to advance certain interests. Also, a lot of push back comes from the government because seemingly low/negative results are interpreted as direct criticism on service delivery - which is not always the case.

Similarly, CSOs are often not trusted because of the belief that a lot of their work is not independent, it is donor-driven and there might be also other hidden interests. The converse is the same for non-state actors; there is a degree of mistrust of data published by the government because of the assumption that data can be manipulated to hide the actual situation of development by the government.

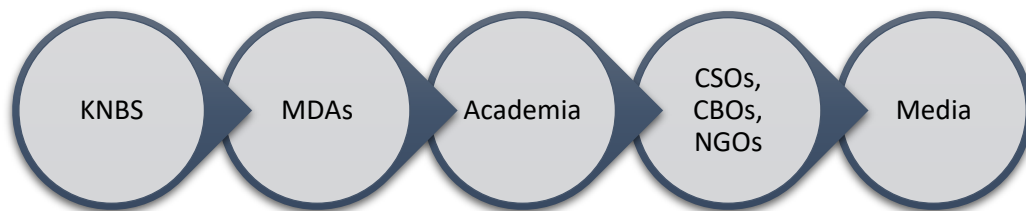


Figure 7: Decreasing levels of trust in data from Left to Right as per interviews

Generally, there is trust in data collected by academia, think tanks, and other international organizations. This is because there is a well-defined process by which data is collected, analysed, verified and disseminated. This study therefore finds that the level of mistrust can be minimized if guidelines and analysis tools are laid out publicly for the citizens and other stakeholders to understand how the data was collected and analysed.

5.9 KNBS Definitions, Guidelines, Norms and Standards

According to KNBS, the population distribution and household distribution are different constructs. In Kenya, 68% of the population lives in rural areas, which means that the average urban household size is smaller than in rural areas⁷⁹. Consequently, the Fifth National Sample Survey and Evaluation Programme (NASSEP V) has a total of 5,360 (2,568 are in urban areas and 2,792 are in rural areas) clusters⁸⁰. This distribution of clusters will allow for further sub-domain analysis and proper weighting of indicators when presenting findings.

The NASSEP V master sample frame was developed from a total of 96,252 Enumerations Areas (EAs) created during the 2009 Kenya Population and Housing Census. Further, administratively, Kenya is divided into forty-seven counties; sub-divided into a number of districts, divisions, locations and finally to sub-locations. During the development of the NASSEP V, each of the 47 counties was stratified into urban and rural strata. Since Nairobi county and Mombasa county have only urban areas, there are a total of 92 sampling strata.

Below are some important definitions from the KNBS Compendium⁸¹ that help to further clarify concepts in sampling.

- 1. Population:** This term applies to the totality of all units of interest in a study or investigation at a given time in a given area.
- 2. Sampling:** The process of obtaining a representative sample from a given population.
- 3. Sampling frame:** A comprehensive list of distinct and distinguishable units within a population from which a sample is selected.
- 4. Sampling Unit:** An element or group of elements of a population, which is convenient for selecting samples. A sampling unit can be a group of persons, households or administrative areas.
- 5. Sampling Design:** The mechanism used to select a sample. The sampling designs are divided into two broad categories: the random and non random sampling design. The types of sampling designs include:
 - a. Simple Random Sampling (SRS):** Under SRS, the desired elements are selected by a chance or probability process and each element in the population has an equal chance of being selected/ chosen.
 - b. Systematic Sampling:** The elements are consecutively numbered and the first unit (k) in the sample is selected randomly. The sample is then selected by taking the kth, (k+s)

⁷⁹ KNBS rationale for sampling; defined in Kenya Integrated Household and Budget Survey 2015-2016. The rural EAs accounts for 61.7 per cent (59,407) of the total, while urban areas represent 38.3 per cent (36,844) of the EAs.

<http://54.213.151.253/nada/index.php/catalog/88/sampling>

⁸⁰ This is what that the KNBS currently operates to conduct household-based surveys throughout Kenya.

⁸¹ <https://www.knbs.or.ke/download/compendium/>

th, $(k + 2s)$ th, ..., etc., where k is not larger than the sampling interval (s). The sampling interval is obtained by dividing the total population by the sample size (n). For example to select a sample of n units, we take a unit at random from the first k units and every k th unit thereafter. If s is 15 and the first unit drawn is number 20, the subsequent units are numbered 35, 50, 65 etc.

- c. Random Sampling with Replacement:** At any draw all the members of the population are given an equal chance of being drawn no matter how often they have already been drawn.
 - d. Cluster Sampling:** This is where the elements are not selected individually but rather in groups, often based on proximity. This allows overcoming the constraints of costs and time associated with a dispersed population. NB: All the elements in the selected group are included in the study.
 - e. Stratified Sampling:** In Stratified Sampling, the population is divided into homogeneous sets or strata with respect to the characteristics being studied and a random sample is selected from each stratum independently.
 - f. Strata:** In stratified sampling the population of N units is first divided into non overlapping subpopulations of N_1, N_2, \dots, N_L units respectively. These subpopulations are referred to as strata.
 - g. Quota Samples:** A sample, usually of human beings, in which each investigator is instructed to collect information from an assigned number of individuals (the quota) but the individuals are left to his personal choice. In practice this choice is severely limited by "controls", e.g. s/he is instructed to secure certain numbers in assigned age groups, equal numbers of the two sexes, certain numbers in particular social classes and so forth.
 - h. Multi-stage Sampling:** A method where selection of the sample is carried out in several stages, i.e. primary units are regarded as a population from which sub-samples of secondary units are selected.
- 6. Survey:** A relatively extensive and comparatively widespread study that uses interviews or self-administered questionnaires as the methods of collecting the data and where sampling techniques are employed to select the respondents or units of interest
- 7. Indicator:** A pointer used specifically to shed light on a variable of interest. This variable may be demographic.

KNBS is also mandated through the Statistics Act of 2006 to establish standards and promote the use of good practices and methods in the production and dissemination of statistical information. This study notes that explicit guidelines, norms and standards are still lacking to guide the process and quality of data collected by non-state actors and county governments. For instance, the fourth schedule of the Kenya Constitution of 2010 and the County

Government Act 2012 (section 104, specifically 105 (e), 106 and 107)⁸² dictate that data gathering is a shared role between the national and county government, though presently, counties are not permitted to undertake certain statistics functions like population census. Nevertheless, there are examples of good practices that have been developed and operationalized by KNBS. Examples include:

1. **A data sharing and dissemination policy⁸³**. This policy is an inward-looking document (that is, it describes how KNBS publishes information; this policy does not outline how non-state actors can publish ‘official’ statistics or data) that provides conditions on how KNBS is to release information as well as the permitted usage of KNBS accessed data and the nature of data files that may be released.
2. **Summarized census guidelines⁸⁴** outline why everyone should participate in the census and how this information helps determine locations for schools, roads, hospitals, and more. Businesses can also use census data to locate supermarkets, shopping centres, new housing and other facilities; amongst others.
3. KNBS has provided inputs into the **Kenya SDG Monitoring framework** adopted in June 2017⁸⁵. This will be discussed further in the next section.
4. **The data quality assessment framework for Balance of Payment Statistics⁸⁶**. This is a comprehensive Data Quality Assessment Framework (DQAF) for Balance of Payments Statistics which also provides very useful guidelines that can extend to other sectoral data.

Table 7: Transferrable principles of data quality borrowed from the Data Quality Assessment Framework for Balance of Payments Statistics and the Principles of Fundamental Statistics.

PREREQUISITES	
Legal and institutional environment	Collecting, processing and disseminating: The statistics act identifies KNBS as the principal agency of the government for collecting, analysing and disseminating statistical data in Kenya and permits KNBS to plan, authorize, coordinate and supervise all official statistical programmes undertaken within the National Statistical System (NSS).

⁸² A county planning unit is charged with driving the county planning framework which integrates economic, physical, social, environmental and spatial planning. Specifically, section 105(e) states that the county planning unit must ensure “... the collection, collation, storage and updating of data and information suitable for the planning process;”

⁸³ <https://www.knbs.or.ke/download/knbs-data-access-and-data-dissemination-policy-2/>

⁸⁴ <https://www.knbs.or.ke/census-frequently-asked-questions-faqs/>

⁸⁵ <http://planning.go.ke/wp-content/uploads/2018/04/SDG-Implementation-plan-2030.pdf>

⁸⁶ Balance of payments statistics record economic transactions between residents and non-residents. Data are collected for all the major balances, inflows and outflows in the current, capital and financial accounts.

	<p>Data sharing: The NSS includes producers and users of statistics under the supervision and coordination of the KNBS and mainly comprises MDAs and county governments.</p> <p>Confidentiality of individual data: KNBS cannot publish, or otherwise make available to any individual or organization, information that would enable the identification of any individual person or entity (Statistics Act - Section 22)</p>
Quality management	<p>KNBS is committed to providing quality statistics and associated services effectively and efficiently. KNBS is committed to a quality management principles - aligned to ISO 9001:2015 international standard - which are implemented by a dedicated quality team.</p> <p>Internal audits are to be carried out on a regular basis through an internal auditor.</p>
INTEGRITY	
Professionalism	<p>Impartiality of statistics: it is important to build a reputation around impartiality, professionalism, scientific approach in the compilation of statistics.</p>
Transparency	<p>Disclosure of terms and conditions for statistical collection, processing, and dissemination: during the conduct of surveys and censuses, respondents are informed on the objectives of these surveys, their rights and obligations for information being provided</p> <p>When there is changes in methodology, the media are notified - through presentations and other notes.</p> <p>Official statistics are published at the same time (to both internal - within government - and external stakeholders). No one is given information before its official release.</p>
Ethical Standards	<p>All KNBS staff must abide by a code of conduct - exercising public trust in a worthy manner and in the best interest of the people of Kenya.</p>
METHODOLOGICAL SOUNDNESS	
Concepts and definitions	<p>This has to do with putting in place the correct research design and analytical framework.</p>
ACCURACY AND RELIABILITY	
Sources of data	<p>Reliable and accurate data sources must be selected. There is also need not to use outdated data.</p>
SERVICEABILITY	

Periodicity and timeliness	To decide whether data will be published quarterly or annually. Once this is decided, the publication should remain consistent.
DATA ACCESSIBILITY	
Accessibility	<p>Presentation: Data is presented in clear charts, maps and tables</p> <p>Dissemination: Reports are published in hard and electronic copy on the KNBS website</p>
Metadata	Metadata is provided alongside the published statistics and is available on the KNBS website
Assistance to users	<p>Mentioning focal persons on the report as well as on the KNBS website</p> <p>Generally availing documents and service catalogues.</p>

All of the above good practices are transferable to the publication of data by non-state actors. They are also embodied in the Principles of Fundamental Statistics (see **Annex C**).



6. CONCLUSIONS

6. CONCLUSIONS

6.1 Data Availability and Quality

County Government Departments of Elgeyo Marakwet, Kisumu and Vihiga collect a lot of data relevant to the work that they are mandated to do. This data is mainly used for planning department activities, project and programme monitoring, guidance in implementing county government interventions and for policy formulation. Depending on the type of data, the collection periodicity ranges from monthly, quarterly and annually depending on resources and skills of the organisation.

CSOs and non-state actors are mostly using the official data for establishing a baseline to help them design programs. Implicitly they know this information is not up-to-date; for example, they can design a program and identify strategic areas to invest using the data but they cannot use it to budget or pick households for commodity distribution. To do this, they may do another count so as to know how many commodities will actually be distributed.

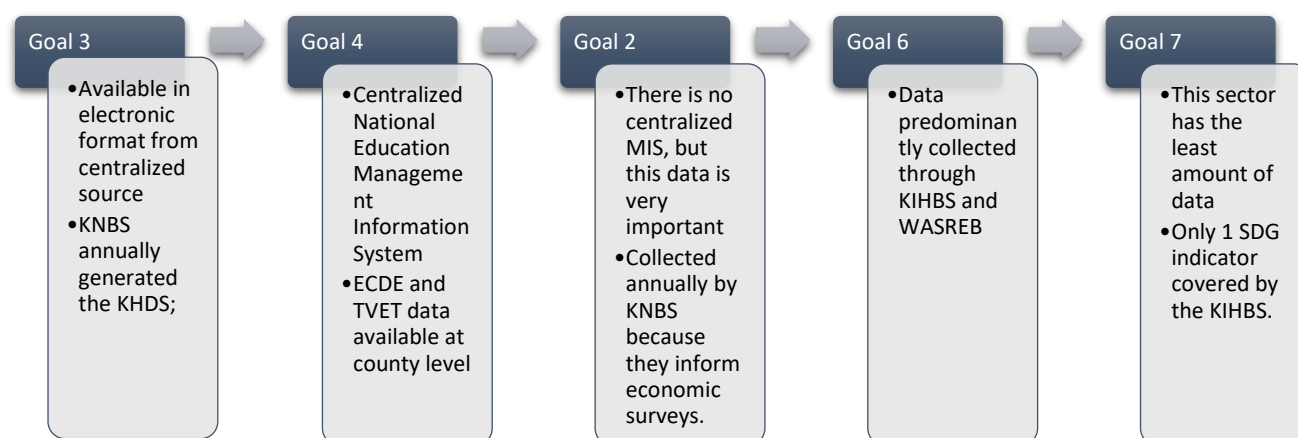


Figure 8: Data availability for the SDGs under study (applicable to both state and non-state actors)

CSOs are also incorporating national level data (from ministries) to their own reports. For example, the figure below appears in the Annual Report 2016/17 prepared by KEWASNET.

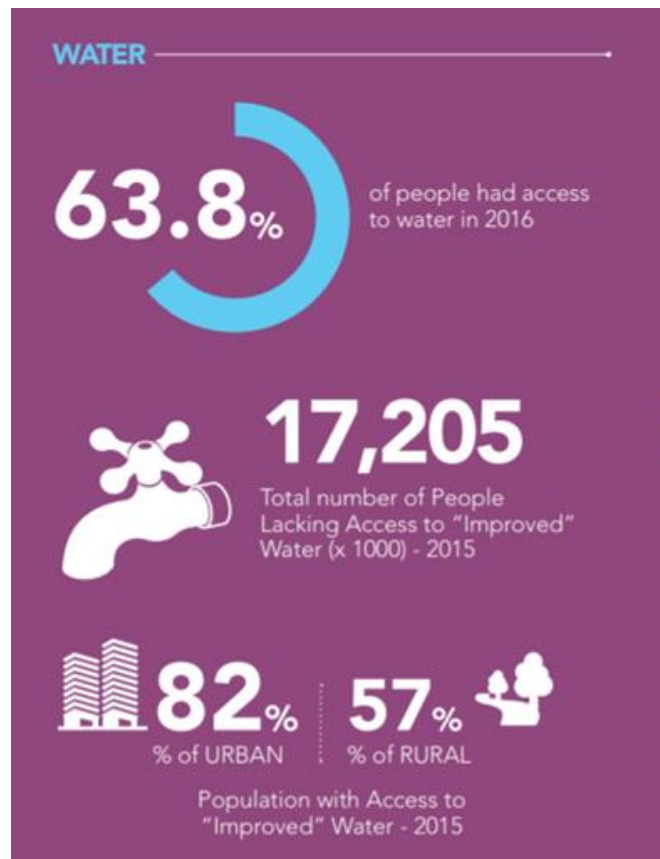


Figure 9: Sample infographic: State of water, KEWASNET 2015/16 Annual Report (source is MoWS, KNBS)

State and non-state actors:

1. Collect data from both primary and secondary sources.
2. Use data to allocate resources to implement interventions.
3. Use data to engage and educate community members. The education may range from how farmers should implement sustainable farming practices, to how citizens should conduct themselves to avoid breakout of diseases such as cholera.
4. Disaggregation is important and the institutions disaggregate data as indicated in section 5.1.1.
5. Workshops and field observations are the more common data validation exercises before publication of data.

Unfortunately, collaboration between different state and non-state stakeholders has been minimal because of mistrust. Government often trusts only its own data and perhaps data from academia, while mistrusting CSOs and the media. On the other hand, CSOs trust academia and do not trust government data. This has negatively affected collaboration as each stakeholder has their own belief on who has authentic and valid data. All stakeholders have data important to the implementation process of the SDGs but only government data is portrayed as official data. The basis for this is that at subnational level, there are gaps/skills

in data production and analysis. KNBS county offices report that even the national level NGOs do not exercise the requisite rigor during sampling, they neither validate data tools (with KNBS) nor results after. But where the methodology is subjected to such rigor, KNBS report that there are no problems with incorporating this into official statistics.

Statistics publication is not a devolved function and therefore, data published by county governments is not considered as official statistics by KNBS; nevertheless, by virtue of the planning function and some limited functions such as data collection, county government departments are producing data which is being cited in the County Statistical Abstracts, GCP and Annual KNBS abstracts. However, the validation process (when the data is submitted by KNBS county statisticians to the national level) takes too long. This time lag (necessitated by the need for KNBS TWGs at the national level to convene and validate data) compared to the requirement that county governments must publish data (as per the County Government Act) to enable public participation shows that there is a lacuna in the current legal framework given that only KNBS is mandated to publish official statistics. The hope is that this can be addressed by the pending county statistics bill.

Further, in the case of non-state actors, as mentioned earlier, KNBS is developing guidelines for methodology to employ in sampling, frequency, actual collection, analysis, and dissemination of data.

County Government Departments of Elgeyo Marakwet, Kisumu and Vihiga collect a lot of data relevant to the work that they are mandated to do. This data is mainly used for planning department activities, project and programme monitoring, guidance in implementing county government interventions and for policy formulation. Depending on the type of data, the collection periodicity ranges from monthly, quarterly and annually depending on resources and skills of the organisation.

6.2 Data and Skills Gaps

County governments and non-state actors working with different SDG programs continue to grapple with the issue of lack of skilled human resource in relation to data collection, processing and dissemination. County departments have to rely on national government or other organisations for statisticians because of lack of in-house statisticians. Departments have a shortage of personnel working in data offices. There is also a shortage of trained field agents involved in the data collection process. In the end facilitators have to rely on locals who are not trained data-collectors which compromises the quality of data collected.

According to county departments and CSOs studied, there are different data sets that are either missing or outdated. In some cases, departments and organisations have relied on the 2009 National Census and projections to estimate data. This becomes problematic in planning and resource allocation as the correct data is not used to inform decision and policy making processes.

For these two instances of skills gaps and data gaps, inadequate financial resources has been cited as the main impediment to filling the data skills gaps. It also means that actors are unable to train staff on data collection. Importantly, it means that when funding for programs is no longer forthcoming or is reduced (and a lot of projects - especially with CSOs - are funded by external donors), either data will not be collected or data component of the project is deprioritized.

6.3 New Ways of Data Collection and Management

From the study, it is evident that government departments and other organisations still use manual processes and tools to collect and store data. In other cases (like in the agriculture sector) in which technology is used to store data, it is not centralized but is stored on staff laptops and flash drives. The dangers of not centralizing data is that it will be in different formats that are difficult to analyze and that can be easily lost.

Also, there are missed opportunities to harness citizen generated data (some of which is already being collected) and big data to improve the timeliness and availability of data.

But there are departments using new ways of data management. Facilities in Elgeyo Marakwet, Kisumu and Vihiga counties upload their data to the DHIS2. In Vihiga, the Department of Environment, Water, Energy and Natural Resources uploads its data to WARIS platform hosted by Water Services Regulatory Board (WASREB). The MoEST has also deployed NEMIS. These are examples of centralized MIS provided by the national government in the context of devolved functions.

6.4 Anonymization of Data

It is evident that during data collection, enumerators (especially from CSOs) have always sought respondents consent to participate in the process. The consent is normally verbal, written or through telephone conversations. The Government often collects data with service delivery in mind; often, people will not get prescribed services if they do not provide requested details. When respondents are reluctant to participate in the data collection process, enumerators have to assure respondents that:

-
1. The data will be put to good use;
 2. That respondent's personal information will remain anonymous; and that
 3. In sharing data, personal information of respondents has been anonymized especially because the data presented is grouped data as opposed to individual data being published.



7. RECOMMENDATIONS

7. RECOMMENDATIONS

7.1 Data Availability and Quality

- 1. Increased collaborations between state and non-state actors**, with KNBS being a focal point, will ensure that: (i) there is no duplication of data collection which will save on resources. (ii) state and non-state actors can complement the already available data. (iii) they make use of different organisations' resources and expertise to harness data collection, analysis, dissemination and storage methods. (iv) indicators need to be standardized at the national level and guidelines for collection and validation developed. This is where CoG and the IGTRC would need to harmonize operational guidelines in consultation with county governments. Also, data from non-state actors working on Goal 6 and Goal 7 can be mainstreamed for monitoring SDGs since CSOs/Academia have the skills and expertise to gather and monitor this data.
- 2. Adequate resourcing** remains a major challenge to effective data collection, analysis and dissemination. Collaboration can help to facilitate crowd funding for data related activities - for both state and non-state actors. Busia County Government is leading the way (they were the best performing in service delivery - with monitoring being a key component⁸⁷) in this; with support of the World Bank under the Kenya Devolution Support Programme (KDSP) in the 2019/2020 budget estimates, there are explicit budget allocations in IFMIS that will be utilized directly under M&E⁸⁸.
- 3. With citizens increasingly reluctant and wary of speaking to data collectors, data education and data literacy** is crucial to informing the respondents importance of the exercise and how continuously supplying the right data will improve services that trickle down to the grassroots through better accountability during implementation of public programmes and projects. Participatory budgeting is concerned with ensuring that resources are invested in the most urgent yet important projects and programs. Through participatory budgeting, with

⁸⁷ <https://cog.go.ke/component/k2/item/97-busia-county-scoops-top-inaugural-performance-for-results-trophy>

⁸⁸ This is a significant change from 2017 when Twenty one counties, including Bungoma, Busia, Elgeyo Marakwet, Embu, Homa Bay, Kajiado, Kericho, Kilifi, Kirinyaga, Kitui, Kwale, Laikipia, Marsabit, Nakuru, Narok, Nyeri, Samburu, Taita Taveta, Turkana, Baringo and Wajir, participated in a voluntary survey undertaken by UNDP geared at assessing M&E capacities at county level. Back then a quarter (23%) of the respondents reported not having any M&E Units/ Divisions in their respective counties - Busia was one of them.

(<https://www.undp.org/content/dam/kenya/docs/Procurement/2017/July/Website%20Change/UNDP%20County%20ME%20Capacity%20Assessment%20Report.pdf>)

time, citizens will get to learn that funding ‘soft’ activities like data collection eventually will improve how decisions are made.

- 4. Data validation and methodology** are critical in avoiding situations where data published is not correctly analysed or has errors. Also, all stakeholders should always publish their methodology so that findings can be subjected to validity⁸⁹ and reliability⁹⁰ tests in order to reduce mistrust. The figure below recommends important skills that are required by different stakeholders.

With **better methodology, the quality of decision making and planning** will certainly improve. In this way, it is time that the definition of ‘officialness’ shifts from ‘who’ is producing the data to the methodology, that is, the ‘how’ data is collected as the definition for official statistics. The figure below builds out the concepts which must be adhered to for production of official statistics:

- a. A quantitative approach that will likely be positivist;** involving a high degree of objectivity (researcher is detached), structured questions and structured analysis. It is best for explanatory research that explains phenomena and focuses on priorities. Data is collected through experiments, surveys, testing, structured content analysis, interviews, and observation. Such research is typically useful to predict what will happen in the future (mostly the ambit of official statistics). Official statistics are by nature practical and measurement tools; they ought to be impartial and made available for public good (see the fundamental principles of statistics - Annex C)
- b. A qualitative approach is more likely to be interpretivist;** the researcher is not detached from the subject; they delve to understand ‘why’ phenomena observed are the way they are. They mostly delve into behaviours. Data is collected through unstructured interviews, observation, and content analysis. Topics include experiences of women, children, youth and populations who are often marginalized in society; such approaches inform social change (mostly the ambit of non-official statistics). For example, baseline data can be obtained before implementation of a social program so as to study the effects of the program by tracking indicators.
- c. Sometimes a combination of qualitative and quantitative methods** is employed by researchers. The focus is still the phenomenon under observation - so that it can be defined better; but this is deemed outside of the scope of this study.

⁸⁹ Validity entails reflections on whether the results are trustworthy and meaningful. Internal validity relates to how well a study has been conducted (its structure) and external validity relates to how applicable the findings are to the real world.

⁹⁰ Reliability entails reflections on whether results can be replicated.

The figure below illustrates that 'how' data is being collected to answer the needs at hand is more important than 'who' does the data collection as long as common/understood guidelines have been adhered to in the design and analysis stages and publication of the methodology.

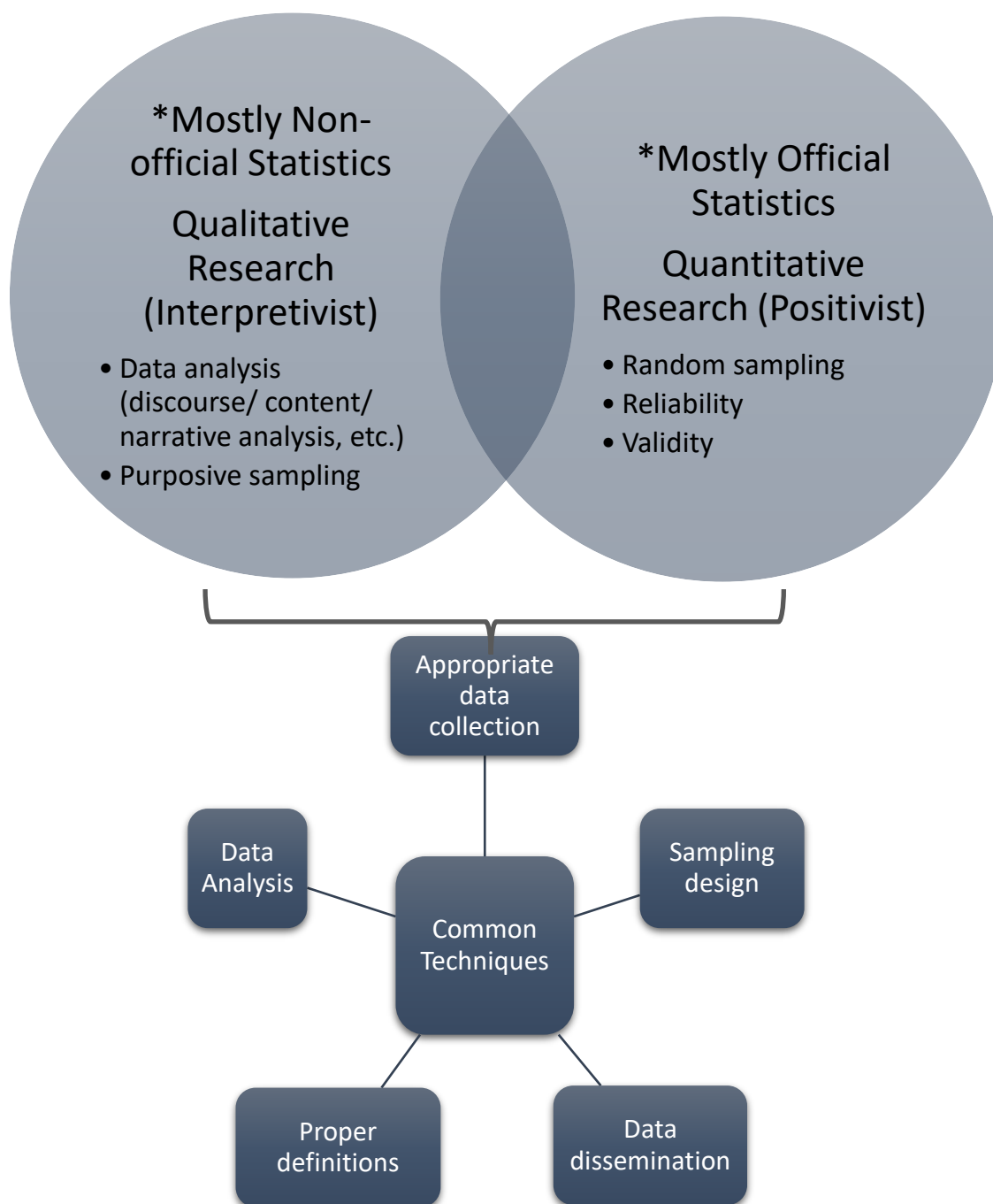


Figure 10: Shifting from 'who' collects data to 'how' data is collected.

There stands to be benefits across board when there is sound methodology (whether within the ambit of official or non-official statistics). The applicable techniques (depending on research paradigm) for data definitions, collection, sampling design, analysis and

dissemination cut across official or non-official statistics. Consequently, a framework like this allows a data producer to know where their initiative would plug in; it also provides guidance in soundly collecting and analysing data to monitor SDG indicators (Tier II and III)⁹¹.

5. There should be distinct **classification of data**:

- a. **Administrative data** – roads, disease burden, enrolment, does not need a lot of validity tests (it can be sampled randomly to ensure that it is representative of the population).
- b. **Survey data** - would require more rigor in terms of validity/reliability tests; KNBS should look at the research instruments; sampling using NASSEP V frame is very important for surveys.

7.2 Data Publication

1. Non-state actors should improve on sharing/publishing findings with multiple stakeholders and focus more on data visualization.
2. **Data publication protocols by state actors (at national and county government level) should be developed** and/or communicated with non-state actors. Unfortunately, sometimes the media (or even CSOs) report highlights and snapshots of the phenomenon without truly understanding the problem and digging deeper to find out data behind stories (or stories behind the data).
3. More **disaggregated data** should be published at the lowest levels (that is, community level). Such data can be used not just for planning resource allocation, but for calling communities to action and to change behaviours since the dealings are directly with the citizens.
4. KNBS has the monopoly on official statistics because of the Statistics Act of 2016. But for certain SDG data indicators, for which they do not traditionally collect data, opportunities are missed and more importantly, populations are left behind. This calls for:
 - a. **Operationalizing guidelines** that explain exactly how data is collected and verified before publication of data from non-state actors.
 - b. Awareness raising on the importance of the **county statistics** bill in order to fast-track its enactment.
 - c. Gender is a legal and constitutional issue and no longer a mainstreaming issue. Given that this dimension is missing in a lot of the data collected, both the operational guidelines and county statistics bill should emphasize this so that the law translates to practice.

Expounding on 4(a) above, drawing from the work that has been done on the county statistical abstracts and GCP, **KNBS should develop and publish operational guidelines** which

⁹¹ <https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/>

provide guidance to non-state actors desiring to undertake surveys or to design public programs. It should encompass key concepts from the NASSEP V frame, the KNBS compendium (of definitions), the KNBS Data Dissemination and Access Policy 2012. Below is a draft process by which data by non-state actors can conform to methodological rigor set forth by KNBS.

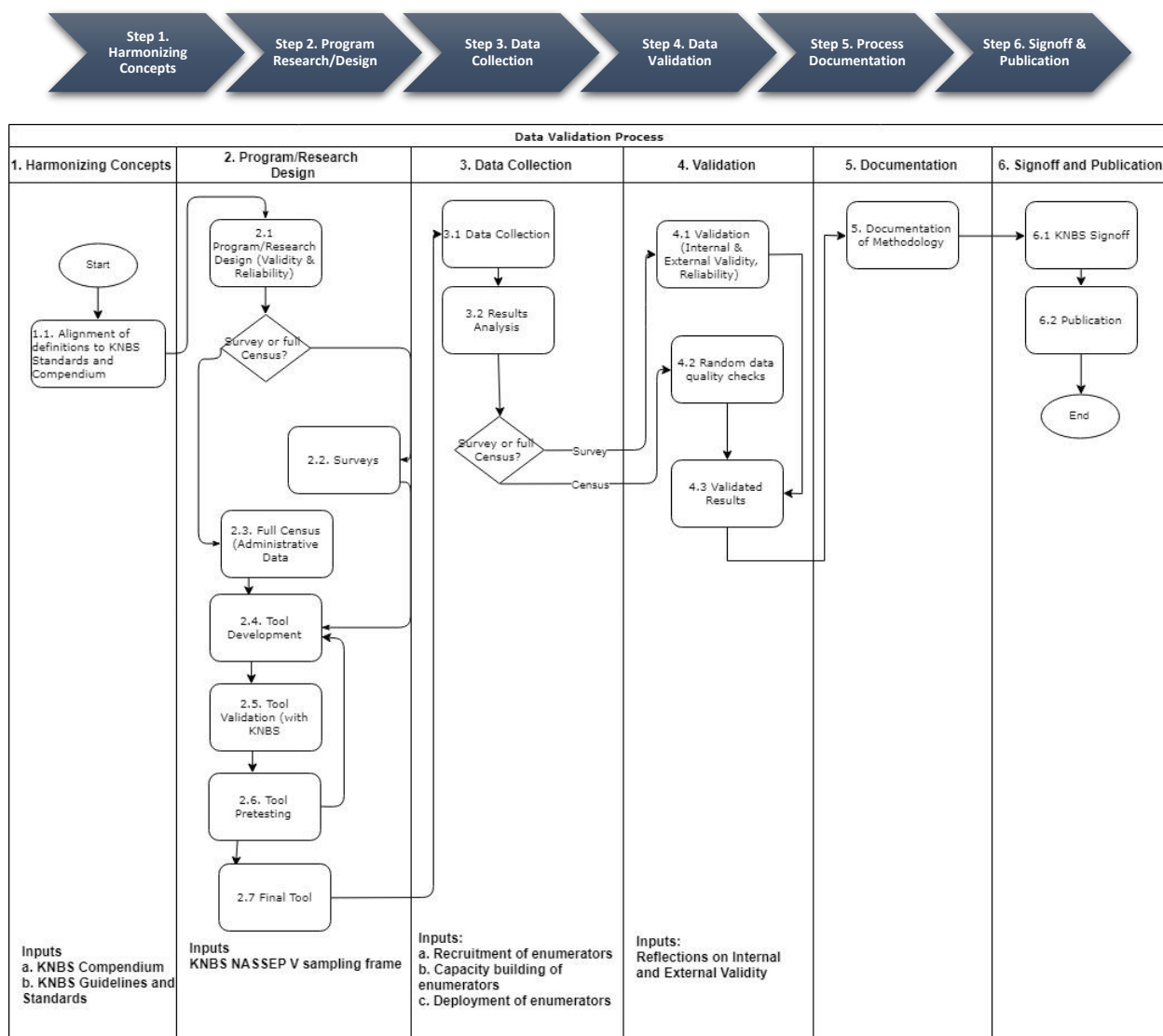


Figure 11: Proposed process (summarized and detailed) for non-state data collection and validation

7.3 Data Skills Gaps

- 1. Champions drawn from the county assembly (specifically, the County Assemblies Forum⁹²) need to be trained on indicators** so as to enrich the debates in the county assembly when discussing policies (that their basis is evidence and data). As one interviewee notes on this issue, it is not possible for a Health Committee of the County Assembly to discuss health matters without first understanding health indicators that are being monitored or the SDGs. This should also include coming up with a sound framework for progress reports for monitoring that do not only consider the financials (budgets and expenditures for M&E activities) but also the actual activities and achievements. For sustainability the training can be anchored in CoG and the Kenya School of Government (KSG).
- The position of this paper is that **there should be deliberate resource allocation to data dissemination to the lowest community levels**. This will involve use of infomediaries and community champions (such as teachers) to simplify findings, data and results; even when a study is targeting complex and abstract policy changes. Citizens need to understand data and policies that ultimately impact them.
- Further-more **community empowerment is needed to understand the importance of data** to policy. Locals are the people who best know what has been implemented (by state or non-state actors) or not. It is key for the success of devolution and for SDGs implementation.
- Strengthening legal and policy framework as relating to human resources and skills for data.

The figure below summarizes important tenets required to strengthen data quality.

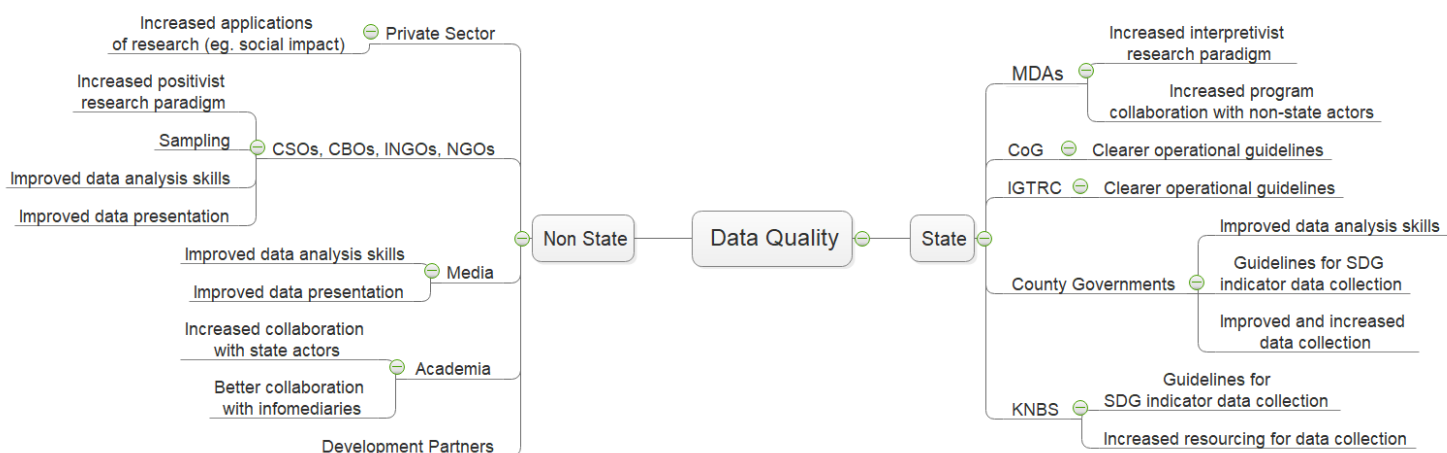


Figure 12: Recommendations on skills strengthening for different stakeholders for improved data quality

⁹² The County Assemblies Forum (CAF) provides leadership and coordination for all 47 County Assemblies in Kenya (<https://www.countyassembliesforum.org/>).

7.4 New Ways of Data Collection and Management

1. A lot of **investment** is required in acquiring new data management methods and tools. The current methods of printed documents for majority of the stakeholders in the county data ecosystem are not sustainable. Having data platforms and other back-up means will prevent loss of data upon retirement or movement of staff who were responsible for handling data.
2. **CGD has become prominent as a new means of data collection** outside the public sector. However mechanisms and frameworks for institutionalizing CGD in the public sector will be required in order to support adoption and investments for proper data storage and analysis in order to leverage it for development. Here, technologies such as SMS for push or pull surveys, smartphone applications can be leveraged to collect any kind of data - including images, GPS locations - which enhances and enriches the level of resultant analytics.
3. **Toll-free lines** (which Kisumu and Elgeyo Marakwet have in place) are avenues by which data can be collected directly and easily from citizens to the government who then find it easier to aggregate priorities and analyse data.
4. Stakeholders should therefore come up with tools where data from different sectors can be accessed and possibly be stored electronically (to ease analysis, storage and sharing) in a centralized location; avoiding flash disks and hard drives and on printed media behind 'closed' access. See example from IFPRI⁹³ that promotes open-access data and analysis on agricultural research investment and capacity in low-and middle-income countries. Another example is CABI which runs a data project called Plantwise that hosts information on plant health issues for access by different users such as environmental managers, researchers, farmers, plant protection officers. Plantwise provides plant health information and services - from diagnostic and management advice to maps of pest locations and customized alerts on pests, among others⁹⁴.
5. **Platforms that bring all stakeholder data together** are important because this helps to avoid duplication during program design, and data collection so that efforts complement each other to secure the missing data. CIMES provides a good opportunity to automate performance monitoring data by departments and which would feed into SDGs - practically, not all the departments can have their own data collection data management information systems. Unfortunately, interviews and observations at national and subnational levels indicate that there is still lack of clarity on actual practices and implementation of both NIMES and CIMES as well as resistance. Hence, for CIMES to be practically adopted, the roll out should consider a bottom-up, participatory design approach that begins to incorporate the views of county departments so that the tools do not just become other national reporting tools, but beneficial for use by the county.

⁹³ <https://www.asti.cgiar.org/kenya/directory>

⁹⁴ <https://www.plantwise.org/KnowledgeBank>

-
- 6. Despite competing interests, stakeholders must strive to use the data which they collect to improve the quality of decision making.** Even if accurate data was collected (timeously) and supported by very expensive infrastructure, but it is never used, a lot of efforts are rendered worthless. One of the reasons for lack of data use is posited in section 5.6 - a lot of citizens might not be digitally literate despite the proliferation of ICT tools. The other reason is resistance - often from state actors - who might interpret non-state produced data as criticism on performance.
 - 7. Structured investments as driven by the National ICT Master plan and the respective County ICT Roadmaps⁹⁵ are needed** so that there is a coordinated effort in ensuring the software and hardware required to support the data infrastructure is achieved - possibly coordinated by CoG and IGTRC)

7.5 Data Collaboration

1. Collaboration amongst CoG and the Intergovernmental Technical Relations Committee (IGTRC) should be defined by law. Right now mandates are vague and there are no accountability mechanisms for roles by these 2 agencies in matters of pertaining to data. **Operating guidelines that coordinate data functions within the executive (at county government level)** but which still embody the original spirit of appreciating the distinctness, interdependence and respect for the mandate and institutions of all levels of government should be developed.
2. **Collaboration between non-state and state actors avoids overlaps in programming** so that duplication does not occur in the same catchment area (and there is no double counting when reporting results).

7.6 Politics of Data

1. Politics seem to play a role that undermines the use of data in implementation of planned interventions. It helps that permanent government employees existing in all counties and these are hired by the county public service board and are not political appointees. Strong advocacy and holding all stakeholders to account will be key in ensuring that politics do not take centre stage in implementing policies.
2. Having well defined and easy to understand data collection, analysis and dissemination methods eliminates instances of mistrust between different stakeholders. For quality, it is important to show how data was validated and verified as another means of reducing instances of mistrust and create collaboration.

⁹⁵ <http://icta.go.ke/county-ict-roadmaps/>

-
3. **Lastly, if one is going to challenge policy makers, one has to use to use rigor and methodology.** Data/research that will impact resource allocation also requires rigor.
 4. Creating awareness to the public on importance of data to demand for quality production and use.
 5. Active public participation in government development process provide incentive for data openness and use in policy and decision making.



.....

8. SUMMARY

.....

8. SUMMARY

This report has mapped the data ecosystem in Elgeyo Marakwet, Kisumu and Vihiga Counties looking at: data collection, use and validation resourcing for data collection, collaboration between state and non-state data producers' data and skills gaps, use of new ways of data management and politics of data. It has also outlined different challenges experienced by data producers when collecting data and when trying to build collaborations between different data producers.

It has been seen that data gaps exist in terms of data unavailability and presence of outdated data that is still used for planning. These gaps are as a result of inadequate financial resources limiting number of skilled personnel employed to participate in data collection, analysis, dissemination and storage. There are gaps as far as collaboration; and centralizing data would go a long way to foster this; through addressing interoperability (within and across both state and non-state actors). The right ICT infrastructure has to be in place for this to take place.

CGD can play a major role in reporting service-related issues to the government. It can be harnessed for additional data related to SDG implementation, for example reporting suspected disease outbreaks such as cholera and reporting cases of Gender Based Violence (GBV). Both state and non-state data producers have experience with this collection method.

Lastly, there would need to be a shift in focus from 'who' collects data to 'how' it has been processed. Methodological rigor, and the development of proper operating guidelines holds the key to a well-defined mechanism for creating the right kind of collaboration and coordination for data for implementation of SDGs.



9. ANNEXES

9. ANNEXES

9.1 Annex A: Interviews Conducted

Table: List stakeholder interviews

Elgeyo Marakwet County Government	<ol style="list-style-type: none"> 1. MR. KANGOGO, COUNTY STATISTICIAN, KNBS 2. MR. JACOB AYIENDA, PLANNING OFFICER, DEPARTMENT OF HEALTH SERVICES & SANITATION. 3. ENG. JAMES KEITTANY, WATER SUPERINTENDENT 4. MR. FELIX KIPNGETICH, PLANNING OFFICER, DEPARTMENT OF FINANCE AND ECONOMIC PLANNING 5. MR. ELISHA TANUI, PLANNING OFFICER, DEPARTMENT OF FINANCE AND ECONOMIC PLANNING 6. MR. JAMES WEKESA, ACTING CHIEF OFFICER, DEPARTMENT OF AGRICULTURE AND IRRIGATION 7. MR. JAMES KUTOYI, MONITORING AND EVALUATION OFFICER, DEPARTMENT OF AGRICULTURE AND IRRIGATION 8. MR. BEN KIBOR, COUNTY CROP DEVELOPMENT OFFICER, DEPARTMENT OF AGRICULTURE AND IRRIGATION
Vihiga County Government	<ol style="list-style-type: none"> 1. DR. PATRICK O. LUTTA, COUNTY DEPARTMENT OF HEALTH 2. MR. VICTOR SEREDE, COUNTY DEPARTMENT OF EDUCATION 3. MR. ANDREW O. OTOMO, COUNTY DEPARTMENT OF ENVIRONMENT, WATER, ENERGY AND NATURAL RESOURCES 4. MR. REUBEN CHUMBA, COUNTY DEPARTMENT OF AGRICULTURE 5. MR. ANDREW LORENG'E, COUNTY DEPARTMENT OF PLANNING 6. MR. FRANKLIN MUGANDA, COUNTY DEPARTMENT OF PLANNING
Kisumu County Government	<ol style="list-style-type: none"> 1. DR. KENNEDY OTIENO, COUNTY DEPARTMENT OF HEALTH - 2. DR. DICKENS ONYANGO, COUNTY DEPARTMENT OF HEALTH - 3. MS. ADAH K. AMEDI, COUNTY DEPARTMENT OF EDUCATION, GENDER AND ICT 4. MR. GEORGE ORUDE, COUNTY DEPARTMENT OF EDUCATION , GENDER AND ICT

	<ol style="list-style-type: none"> 5. MR. FREDRICK ONYANGO, COUNTY DEPARTMENT OF EDUCATION, GENDER AND ICT 6. MS. ANNE KOMBIJA, ACTING DIRECTOR, DEPARTMENT OF WATER, ENVIRONMENT, NATURAL RESOURCES AND CLIMATE CHANGE 7. MR. SYLVESTRE OKETCH, DIRECTOR OF AGRICULTURE, DEPARTMENT OF AGRICULTURE, LIVESTOCK, FISHERIES AND IRRIGATION
KISUMU COUNTY ASSEMBLY	<ol style="list-style-type: none"> 1. HON. SALLY OKUDO (MEMBER OF THE COUNTY ASSEMBLY - CENTRAL SEME) 2. LUCAS OKUMU
National NGOs, CSOs	
AMREF	DR. ELIZABETH WALA, PROGRAM DIRECTOR, HEALTH SYSTEMS
Twaweza	DR. EMMANUEL MANYASA, UWEZO KENYA COUNTRY MANAGER
CoG	MR. KEN OLUOCH
Access Coalition	<ol style="list-style-type: none"> 1. MS. EDITH KARIMI, COMMUNICATIONS OFFICER 2. MS. JACQUELINE KIMEU, REGIONAL COORDINATOR
KNBS	MS. LEAH WAMBURU,
Kenya Water and Sanitation Civil Society Network (KEWASNET)	MR. VINCENT OUMA, HEAD OF PROGRAMS
CAB International (CABI)	DR. ELIZABETH NAMBIRO, PLANTWISE REGIONAL COORDINATOR, AFRICA
Kisumu CSOs	
ABLED DIFFERENTLY - PWD	MS. EUNICE ATIENO
KISUMU ACTION DISABILITY DEVELOPMENT NETWORK - PWD (KADDNET - PWD)	MR. NICK ONDWAT
SOCIAL COUNSELLOR FOR THE DEAF	MR. DAVID ONGOWO
GRASSROOTS TRUST	MS. MARTHA MUGO
TEAM	MR. GEORGE COLLINS OWUOR
NATIONAL TAXPAYERS ASSOCIATION	MR. NICHOLAS ODONGO
NETWORK FOR ADOLESCENT AND YOUTH OF AFRICA (NAYA) KENYA	MR. ABDALLA DAVID
NATIONAL TAXPAYERS ASSOCIATION	MR. RONALD OMONDI

COMMUNITY EMPOWERMENT MEDIA INITIATIVE (CEMI)	MS. LILIAN NYAIDHO
NATIONAL MUSLIMS LEADERS FORUM (NAMLEF)	1. MR. HASSAN HUSSEIN 2. MR. SHALMAT KASSIM
BISHOP ABIERO SECONDARY SCHOOL	1. MS. KHAHOYA VALARY 2. MR. RODNEY OCHIENG 3. MS. EVELYN WANYANGA
NYAKACH KWE WOMEN FORUM FOR NETWORKING (NYAKWEFNET)	MS. CAREN ODERO
KIBOS SCHOOL FOR THE VISUALLY IMPAIRED	1. MR. EMMANUEL BIWOTT 2. MS. CAREN OCHIENG 3. MS. EUNICE AKEYO 4. MR. EDWARD MUTESA
YOUTH ADVISORY COUNCIL	MR. MAURICE JUMA
ADWOG	MS. MERITA ANYANGO
TINADA YOUTH ORGANIZATION	MR. DOUGLAS OTIENO
WEST KENYA DEAF	MS. GAZANZI GETRUDE
MASENO UNIVERSITY	MS. KAREN NYANGARA
	Vihiga CSOs
HAMISI CIVIL SOCIETY ORGANISATION NETWORK - VIHIGA	MR. STEPHEN BULEEMI (also Chair of Vihiga CSOs Network)

9.2 Annex B: Some of the surveys undertaken by KNBS

KNBS undertakes two broad types of surveys, namely, household-based surveys and establishment based surveys. These include – but are not limited to – the following:

1. Kenya Population and Housing Census
2. Kenya Integrated Household Budget Survey
3. Kenya Demographic and Health Survey
4. Malaria Indicator Survey
5. Disability Survey
6. Kenya AIDS Indicator Survey
7. Multiple Indicator Cluster Survey
8. Consumer Price Index
9. Producer Price Index
10. Labour Force Survey

-
11. Micro and Small Enterprise Survey
 12. Child labour Survey
 13. Agriculture and Livestock Survey
 14. Time Use Survey
 15. Foreign Investment Survey
 16. Integrated Survey of Enterprises
 17. Informal Cross Border Trade Survey
 18. Survey of villas, cottages and campsites
 19. Financial Access Survey
 20. Inbound Outbound Survey
 21. Survey on Expansion of Coverage of Financial Institutions
 22. Annual Survey on Industrial Production
 23. Building and Construction Survey
 24. Survey of Quarrying and Mining
 25. Census of Industrial Production

9.3 Annex C: Fundamental Principles of Official Statistics⁹⁶ (revised, 213)

1. **Principle 1.** Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.
2. **Principle 2.** To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional
3. **Principle 3.** To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.
4. **Principle 4.** The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.
5. **Principle 5.** Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

⁹⁶ Source: <http://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>

- 6. Principle 6.** Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.
- 7. Principle 7.** The laws, regulations and measures under which the statistical systems operate are to be made public.
- 8. Principle 8.** Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.
- 9. Principle 9.** The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.
- 10. Principle 10.** Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

9.4 Annex D: Data Collection Tool

SDGs Data Ecosystem Mapping - Targeting Counties and other Subnational Actors

This section will be used to provide the respondent with a background to the data ecosystem mapping.

Background to the Tool

As part of the Monitoring and Review of SDGs in Kenya Programme, GIZ Kenya has commissioned a comprehensive mapping of the SDGs data ecosystem. The objectives of this assignment from our understanding is two-fold, we'll need to:

- 1.** *Conduct a data ecosystem mapping for Kenya. This will include; data collected by official and non-official data producers, tools and methodologies used for data collection and the main data producers (and users) for the data. In order to allow for an in-depth analysis, this data ecosystem mapping will be limited to a number of selected SDGs indicators to be defined by key stakeholders. It will focus both on the national level in Kenya and on the county level (specifically Kisumu, Vihiga and Elgeyo-Marakwet Counties).*
- 2.** *Develop recommendations on how to build an integrated multi-stakeholder approach to SDGs monitoring in Kenya based on the mapping exercise.*

Purpose

To monitor the implementation of the SDGs, high quality data is needed. Consequently, the German Federal Ministry for Economic Cooperation and Development (BMZ) commissioned GIZ to implement a component that is concerned with monitoring and review of the SDGs.

This aims to build on the work of the Government of Kenya (GoK) and other non-state actors to monitor and track performance of SDGs. The component objective is to strengthen the cooperation between governmental and non-governmental data communities that contribute to monitoring the implementation of SDGs in Kenya.

SDG Goals in focus (tick applicable)
1. Goal 2 - Zero Hunger
2. Goal 3 - Good health and well-being.
3. Goal 4 - Quality education.
4. Goal 6 - Clean water and sanitation
5. Goal 7 - Affordable and clean energy

QUESTIONS

I. Data availability and quality.

a. About data and data types

1. What are your focal program areas (or what are the key sectors of focus) ? Highlight SDGs and provide specific activities undertaken under each.

SDG	Highlight of Activities
Goal 2	
Goal 3	
Goal 4	
Goal 6	

Goal 7	
---------------	--

2. What data is being collected for the particular SDG(s) under which your work falls?

SDG	How easy is it to access the data?	Data source (Primary, Secondary)	How collected (Online/offline, Key Informant Interviews, Focus Group Discussions, Observation, etc.)	Examples of data
Goal 2				
Goal 3				
Goal 4				
Goal 6				
Goal 7				

3. What type of data do you collect? (baselines, mid-term? evaluation? Qualitative? Quantitative?)

4. What purposes do you use your data for? Is it to (tick one or more and give examples or provide sample case studies/reports demonstrating 'from data to results')

- a. Educate: _____
- b. Policy or Project Design: _____
- c. Gather baseline data: _____
- d. Help manage services and interventions: _____
- e. Define and refine policy: _____
- f. Engage with communities: _____
- g. Inform research: _____
- h. Monitor performance: _____
- i. Make regulatory decisions: _____

5. As part of your data collection, do you have a file with additional descriptions for sources of data, categorizations, definitions of terms, etc.? (Y/N)

- a. If Yes, provide some examples: _____
- b. What tools (if any) do you use for: _____

- c. Data collection: _____
- d. Data processing: _____
- e. Data dissemination: _____
- f. Data storage: _____

- 6. Is your data disaggregated by sex, education, age, disability or other relevant contexts? If so, explain
- 7. a. Does the data look correct in terms of accuracy, reliability and timeliness? (Y/N)
b. If Yes above, Please explain _____
- 8. Do you validate the data collected? (Y/N)
a. If Yes, How is the process of validation
b. How do we shift the focus from WHO collects the data to the QUALITY of data produced?

b. Resourcing for data collection

- 9. How do you finance your data collection process (if at all)?
a. Is it externally funded (also if a project is funded externally)
b. or you budget for it? (private sector, govt)
- 10. How often do you collect the data? Why?

Sectors	How do you sample? (ie. repeat same population or generate new samples) Is it purposive or not?	How often?
Goal 2		
Goal 3		
Goal 4		
Goal 6		

Goal 7		
---------------	--	--

11. If resources were not an issue, how would you improve your current data collection processes?

C. COLLABORATION

12. a. Do you collaborate with KNBS on data issues? (Yes/No)

b. If Yes, or Not, please indicate the reason: _____

13. Does your organization participate in any national/regional data platforms?

14. Who supplies data to you?

a. *Directly?*

b. *Indirectly?*

15. Who uses your data (or whom do you supply data)? (Hint: think about the data ecosystem and roles to these might be playing - data producers, collectors, analysis, publishers, users)?

a. *Directly?*

b. *Indirectly?*

16. Collaboration

a. Are there opportunities for new collaboration? (Highlight not more than 5 for each of these roles: data producers, collectors, analysis, publishers, users)

b. If Yes, what additional information/requirements/skills about the current organizations or people would be helpful to make this happen?

d. Data publication

17. Do you publish data?

a. If Yes,

b. How available/accessible is your data to other data users?

-
- c. What is the process of accessing your data? (is it available for public, do you pay? email? etc.)
 - d. If No, why?

18. Is data shared reciprocally (across departments and across organizations)?

19. How is data released?

- a. workshops
- b. publications
- c. website downloads
- d. APIs, portals...
- e. Other
- f. (Not at all)

20. What challenges do you face in the process of data dissemination?

21. Do you get feedback on published data and is there a feedback mechanisms for published data?

- a. If Yes, how is such feedback considered in future data collection plans⁹⁷?

II. Data and skills gaps.

1. Missing data (Note to the interviewer: Note to the interviewer: the sector the organisation working in may be interlinked with other sectors where data availability is limited)

- a. What data is missing to enable you do your work?
- b. Why has this arisen?
- c. What can be done to fill this gap?

2. In your field of work, what personnel skills are missing and should be prioritized?

- a. collect
- b. analyze
- c. visualize
- d. Disseminate

⁹⁷ Because the point here should be, when there is a feedback mechanism, it allows follow up (with corrections/additions)- it is a mechanism of quality assurance, in a way. So, let us be less interested in the process and more on substance.

-
3. Do you have any guidelines (and manuals) which you use to train users? eg. how to clean data, to key-in data digitally, etc. _____

III. Use of new ways of data management

1. Are you using any ICT techniques to collect, process, store or disseminate information?
 - a. If Yes, what technologies are you using?
 - b. Are they improving efficiency and reducing costs or have increased overheads?
2. Are you using mobile technology or thinking of using mobile technologies in any way (collection or use)?
 - a. How far (or not) are you in this process?
3. Are there any existing initiatives using (or planning to use) citizen generated data? Why? What is the reason? (Note to the interviewer: probe why there is interest increasingly to generate or use Citizen Generated Data (CGD). Is it because of timeliness? context specific? etc.⁹⁸)
4.
 - a. Is your organization involved in CGD?
 - b. Why are you not able to rely on conventional data?
 - c. For what purpose do you collect CGD??
5. What is your understanding of big data⁹⁹?
6. Are you using big data in any way? If so, what are you currently doing (or plan to do) with big data?

⁹⁸ Categories to analyze this question:

1. Production: sample surveys, household surveys, (understanding citizens experiences in public facilities, services, etc.)
2. Defining: problem scoping. Citizens can visit and describe sites to collect new data or enhance existing information.
3. Dissemination: Approaches such as community scorecards organise group deliberation by facilitating focus group meetings with different groups of people (usually split by sex, age and other relevant criteria). The goal is to collectively define assessment criteria for public services based on people's perceptions of the most critical problems
4. Analyzing: Compiling is often a necessary step towards other analytical tasks that are not possible with individual datasets, be it data definition in the beginning of a project, pattern recognition, cross-verification or others.
5. Enriching of existing data by complementing gaps that are missing. For instance, citizen groups could want to ensure the reliability and accuracy of their data by comparing it against official data collections or prediction models.

⁹⁹ Big Data has been described as data sets that are impossible to store and process using common software tools, regardless of the computing power or the physical storage at hand

IV. Anonymization of data

1. How do you obtain consent from respondents?
 - a. Verbal
 - b. Written
 - c. Other
2. When you share data with other organisations/departments, what procedures do you follow to ensure anonymity/or protect sensitive information?
3. What techniques do you use to assure respondents of confidentiality? And that their responses will not be used in any other way?

V. Politics of data: engaging is not a technical question but a political obligation

1. What are some of the risks encountered data collection, analysis, management and dissemination?
2. Which organization do you trust the most when they publish data? (Rating of 1 to 5; 1 least trust and 5 most trusted)
 - a. *Government*
 - b. *CSOs*
 - c. *Media*
 - d. *Academia*
 - e. *Private sector*
3.
 - a. Do you have experience where you witnessed conflicting information between reality on the ground and data produced by sector actors (either regional, national, subnational)? Y/N
 - b. If Yes above, do you have suggestions on how to address this?



german
cooperation
DEUTSCHE ZUSAMMENARBEIT

giz Deutsche Gesellschaft
für Internationale
Zusammenarbeit (GIZ) GmbH



SDGs KENYA FORUM
For Sustainable Development

SUBMITTED BY: SDGS KENYA FORUM