

DOI: <https://doi.org/10.57125/FED.2025.12.04>

How to cite: Onsomu, E., Nyakundi, G., & Kathenge, J. (2025). The role of data governance in advancing quality education: Insights from Kenya's education sector. *Futurity Education*, 5(4), 52–81. <https://doi.org/10.57125/FED.2025.12.04>

The Role of Data Governance in Advancing Quality Education: Insights from Kenya's Education Sector

Eldah Onsomu

PhD, Executive Director, The Kenya Institute for Public Policy Research and Analysis, 2nd Floor, Bishops Garden Towers, Bishops Road, Nairobi, Kenya, <https://orcid.org/0000-0002-8364-8333>

Gideon Nyakundi*

MSc., Senior Statistician, The Kenya Institute for Public Policy Research and Analysis, 2nd Floor, Bishops Garden Towers, Bishops Road, Nairobi, Kenya, <https://orcid.org/0009-0001-1803-2014>

Japheth Kathenge

MSc., Research Assistant, The Kenya Institute for Public Policy Research and Analysis, 2nd Floor, Bishops Garden Towers, Bishops Road, Nairobi, Kenya

***Corresponding authors:** gnyakundi@kippra.or.ke.

Received: July 3, 2025 | **Accepted:** September 22, 2025 | **Available online:** November 28, 2025

Abstract: Kenya's youthful demographic offers substantial potential for socio-economic transformation; however, weaknesses in education data management continue to limit progress toward quality, equity, and effective human capital development. This study analyses the state of data governance in Kenya's education system and assesses how governance structures, processes, and tools shape educational outcomes. Using the United Nations Department of Economic and Social Affairs data governance framework, a mixed-methods design combined a policy desk review with 193 key informant interviews across 36 countries. Five governance pillars — roles and responsibilities, privacy standards,

processes and procedures, policies and guidelines, and tools and practices — were evaluated across six operational dimensions to produce a national Data Governance Index (DGI). As the first systematic application of a DGI to the education sector in Sub-Saharan Africa, the study offers a benchmark for linking data governance to education quality and human capital formation. Kenya achieved a moderate DGI score of 0.58, with notable strengths in data quality (0.69), accessibility (0.65), and timeliness (0.59). Basic education institutions using the Kenya National Integrated Education Management System outperformed Technical and Vocational Education and Training and Early Childhood Development Education systems, which exhibited significant gaps in metadata documentation, security, and stewardship. County disparities were substantial: Kiambu and Narok recorded DGI scores above 0.80, whereas Isiolo and Tharaka Nithi recorded scores below 0.40. A modest positive correlation between DGI and the Human Development Index — particularly in data quality ($r = 0.24$) — underscores the role of governance in advancing human development. The study advances Human Capital Theory by positioning education data systems as mediators between policy and outcomes. Key recommendations include standardising metadata, automating validation, strengthening county-level security, and institutionalising stewardship roles. The findings provide a roadmap for Kenya and other developing countries seeking to strengthen education reform through robust data governance.

Keywords: Data Governance, Human Capital Development, Education Transformation, Digital Development, Education for Sustainable Development Goals.

Introduction

Education is universally recognised as a key driver of economic development, social equity, and personal empowerment. In the 21st century, data has become an indispensable tool in transforming education systems worldwide. From planning and resource allocation to learner tracking and policy evaluation, data empowers governments to respond to challenges and craft evidence-based solutions (McCarthy et al., 2023; Wang et al., 2023). As the Organisation for Economic Co-operation and Development (OECD) notes, data governance refers to the technical, regulatory, policy, and institutional arrangements that manage the data lifecycle, from creation to deletion, across national and organisational boundaries (OECD, 2022).

In advanced economies, the integration of data governance into education systems has enabled more targeted interventions, improved monitoring of student performance, and strengthened accountability. Countries like Estonia, New Zealand, and Canada have implemented robust data governance frameworks, ensuring interoperability of education systems and protecting student privacy rights while maximising data value (Watson, 2019; Jim & Chang, 2018). These frameworks are not only technical blueprints but also reflect a culture of transparency, ethical standards, and data stewardship.

In Sub-Saharan Africa, the urgency for effective data governance in education is amplified by demographic and socioeconomic realities (Odhiambo, 2024). The region is home to the youngest population in the world, with over 60% under the age of 25 (World Economic Forum, 2023). This "youth bulge" presents both an opportunity for development and a formidable challenge. If properly educated and skilled, this demographic could drive innovation, productivity, and sustainable development (Osei-Kusi et al., 2024). However, as the African Union's Continental Education Strategy for Africa (CESA 2016–2025) acknowledges, education systems across the region often suffer from inefficiencies that stem from poor data management.

Despite progress, systemic weaknesses persist:

i) Education data is often collected by various actors using incompatible systems (Briggs & Cafaro, 2021). This results in siloed information that cannot be aggregated or compared across institutions or jurisdictions. For example, A UNESCO IIEP comparative analysis of six Sub-Saharan countries (Gambia, Ghana, Guinea, Namibia, Senegal, Zambia) found that learning assessment and EMIS data were frequently underutilised in national planning processes, due to weak coordination, limited dissemination, and capacity gaps (Raudonyte & Foimafisi, 2022).

ii) Outdated or incomplete data continues to hamper strategic planning. Materu (2007) highlighted how inaccuracies in school enrollment figures and teacher deployment in several African countries led to misaligned budgeting and resource allocation.

iii) Disparities in digital infrastructure and institutional capacity mean that many regions lack the tools or expertise to access and analyse education data, leading to inequities in planning and response strategies (Tchamyou et al., 2019; Ojiemhenkele et al., 2024).

i) Perhaps most critically, data is often not used effectively in decision-making. This is partly due to weak data infrastructure, lack of analytical capacity, and low levels of data literacy among education administrators. As a result, many education policies and interventions are not aligned with real needs on the ground. Wamutoro et al. (2022) observed that education officers in Uasin Gishu County rarely used EMIS data for strategic decision-making, often due to limited technical capacity and insufficient integration of data use into institutional processes, resulting in policies that may not adequately reflect ground-level needs or promote equity.

Without a coherent governance framework to ensure quality, secure, and ethical data practices, the potential for education data to drive inclusive and responsive development in Sub-Saharan Africa remains largely untapped.

Kenya has made commendable progress in expanding access to education. According to the Kenya National Bureau of Statistics (KNBS), there were 49,033 registered learning institutions as of 2023, serving approximately 15 million learners from pre-primary to tertiary levels (KNBS, 2024). This growth has been accompanied by digitisation efforts, including the National Education Management Information System (NEMIS), biometric student attendance tracking, and digital curriculum delivery platforms.

However, significant data governance gaps remain. These challenges mirror those observed in other sectors, such as agriculture, where the Ministry of Agriculture, Livestock, Fisheries, and Cooperatives (MoALFC) implemented a national data governance framework to address fragmentation and ensure consistency. The success of that framework offers a valuable model for the education sector.

By adopting a comprehensive data governance approach grounded in international best practices and adapted to local realities, Kenya can transform its education system into a more responsive, accountable, and inclusive engine for national development. Such a framework would not only support the realisation of the Kenya Vision 2030 and the Sustainable Development Goals (SDGs) but also ensure that every learner in Kenya is counted, protected, and empowered through the responsible use of data.

Human capital plays a pivotal role in national development. Among its key determinants, education stands out as the most direct and measurable contributor. According to Human Capital Theory, education enhances individuals' productivity, earning capacity, and capacity for innovation, thereby generating positive externalities for society and the economy as a whole (Becker, 2009; Schultz, 1961). In Kenya, where over 70% of the population is under 35 years of age, education is both a strategic investment and a critical foundation for long-term socioeconomic transformation.

Yet the effectiveness of education as a tool for human capital development is increasingly mediated by how well education data is governed. Effective data governance ensures that policies, resource allocation, and reforms are evidence-based, equitable, and responsive. Without reliable, timely, and secure data, education systems operate in the dark, unable to accurately diagnose learning challenges, monitor outcomes, or evaluate policy interventions. Data governance in education, therefore, becomes not a peripheral administrative task, but a central pillar of national human capital strategy.

In this context, investing in robust education data systems is imperative. Kenya's efforts to digitise educational data through platforms such as the National Education Management Information System (NEMIS) represent progress. However, persistent challenges related to data fragmentation, quality control, limited metadata documentation, and weak data stewardship continue to constrain the value of such systems. These issues undermine Kenya's potential to improve its Human Capital Index (HCI), particularly regarding learning-adjusted years of schooling.

Theoretical and empirical evidence strongly support the case for improved data governance. Becker (2009) underscores that education is not merely a consumption good but a productive investment with measurable returns. Schultz (1961) reinforces this by framing education as a public good with transformative potential for nations. Heckman (2008) highlights early childhood education as a particularly high-yield investment in breaking cycles of poverty and improving long-term health and labour outcomes. Chetty et al. (2014) demonstrate that the accurate assessment of teacher effectiveness through value-added models can predict long-term student earnings, affirming the role of micro-level data governance in macro-level economic outcomes.

At the macro level, Hanushek and Woessmann (2022) present compelling evidence that education quality is the principal driver of long-term economic growth. Their concept of "knowledge capital" emphasises the importance of not just the quantity, but the effectiveness with which learning is delivered and assessed. Similarly, the World Bank (2018) stresses that learning outcomes, not enrolment rates alone, determine the real return on investment in education. These findings underscore the urgency of moving from enrolment-driven to learning-driven education systems, a transition that depends on high-quality data governance. Moreover, the link between education data governance and policy effectiveness is well established. Barrera-Orsorio and Linden (2009) find that countries that invest in education data systems and use them systematically report improved educational quality, access, and equity. In Kenya, leveraging such insights can enhance decision-making in areas such as curriculum development, teacher deployment, infrastructure planning, and equity targeting.

In sum, education data governance serves as both an enabler and an amplifier of human capital development. Effective data governance ensures that investments in education are strategic, impactful, and inclusive. For Kenya to realise the full potential of its education sector outcomes in advancing human capital, effective data governance must be institutionalised as a core element of policy design and implementation. This includes operationalising a national education data governance framework with clear roles, standards, and safeguards, supported by training, funding, and institutional accountability.

In light of the gaps and opportunities outlined, this study aimed to critically assess the state of data governance in Kenya's education sector. The study aimed to understand current practices, identify existing gaps, and explore the implications of improved data governance on education quality and human capital development. Specifically, the research questions include:

1. What is the current status of education data governance in Kenya's education sector across the six operational dimensions?
2. How is the Data Governance Index (DGI) associated with human development outcomes (HDI) across counties in Kenya?

The hypotheses tested include:

1. **H1_a:** There is a positive correlation between the overall Data Governance Index (DGI) and county-level Human Development Index (HDI) scores in Kenya.
2. **H1_b:** Among the six governance dimensions, data quality and metadata documentation have the strongest positive correlations with HDI.

Through this analysis, the study contributes to ongoing efforts to enhance transparency, accountability, and equity in the education system through responsible and effective data use.

Methodology

To explore the current state of data governance in Kenya's education sector, this study employed a combination of qualitative and quantitative methodologies with three primary approaches:

Related desk review

A comprehensive desktop review examined national and international literature on data governance policies, strategies, and frameworks. Key documents reviewed included the Data Protection Act (2019), the Basic Education Act (2013), the National Education Sector Strategic Plan (NESSP 2018–2022), and sector-specific documents from ministries and agencies such as the Ministry of Education (MoE), Teachers Service Commission (TSC), and Kenya National Examinations Council (KNEC). In addition, the study drew insights from global and regional frameworks, including the George Washington University (GWU) Data Governance Framework, the Continental Education Strategy for Africa (CESA), and the African Union Digital Transformation Strategy. These resources provided context on data governance principles, institutional roles, privacy and security standards, and emerging best practices.

Key informant interviews (KIIs)

Structured interviews were conducted with selected informants across the education sector to gather in-depth perspectives on current data governance practices, challenges, and opportunities. Participants included Ministry of Education officials, county education directors, quality assurance officers, data protection officers, NEMIS administrators, and education data specialists. The participants were selected from both the national government and county government institutions. The interviews provided insights into how data is managed, shared, and secured in practice, highlighting both the strengths and the barriers faced at operational and policy levels. These insights were triangulated with findings from the desktop review to ensure a comprehensive understanding of the context. A Data Governance Index (DGI) was computed from data and information collected from various stakeholders across counties to indicate the state of data governance in Kenya. The data utilised to compute the national and county-level indices were based on the Institute's 2024 assessments.

The five foundational pillars of data governance: Roles and Responsibilities, Privacy Standards, Policies, Tools and Practices, and Processes and Procedures, were assessed using six key operational dimensions. The pillars and operational dimensions were adopted from the UN Department of Economic and Social Affairs (UNDESA) Data Governance Framework (2025). The dimensions provide a measurable lens for understanding the maturity, gaps, and strengths of existing governance structures. The data collection tool is provided in the Annexe. Table 1 outlines the dimensions and sample statements by stakeholders. The respondents included the heads of the early childhood education and vocational education in the county governments, the county and sub-county quality assurance officers or the county director of education in the Ministry of Education (National Government), or their representatives. A total of 193 key informants from both the national (27%) and county governments (73%) responded to the tool, representing 36 of the 47 counties in Kenya. The sample size of 193 key informants across 36 counties was

determined to ensure adequate statistical power for correlation analysis and meaningful variation across regions. Following common guidelines for correlation studies (Lovakov & Agadullina, 2021), a minimum sample of 152 was required to detect medium effect sizes ($r = 0.20-0.30$) at a 95% confidence level with 95% power. The achieved sample ($n=193$) exceeded this threshold, strengthening the robustness of quantitative analyses. Reliability of the structured tool was assessed using Cronbach's alpha for each dimension. Coefficients ranged from 0.72 (metadata documentation) to 0.84 (data quality), all above the 0.70 threshold recommended for internal consistency (Malkewitz et al., 2023). The interview guide was reviewed by three education data governance experts for content validity and piloted with five officials not included in the main study. Feedback was used to refine question wording, sequence, and clarity, ensuring alignment with the study's conceptual framework.

Table 1

Data Governance Dimension

Data Governance Dimension	Sample Statements by stakeholder
Data Quality	Regular training is conducted to maintain and improve data quality standards
Data Security	Measures (e.g. Data encryption and access control) are in place to ensure the confidentiality of sensitive data.
Data Access	Data in the sector is readily available to authorized users
Data Timeliness	Data in the sector is collected and reported in a timely promptly
Data Stewardship/ Leadership	Data roles and responsibilities are well-defined and understood.
Metadata Documentation and organization	Metadata documentation is comprehensive and up-to-date

Adopted from the UNDESA's Data Governance Framework Elements (2025)

To ensure data quality and minimise bias, interviews were conducted by trained enumerators using standardised protocols. Triangulation was applied by cross-verifying interview insights with desk review findings and county-level education statistics (Schlunegger et al., 2024). Anonymity protocols reduced social desirability bias, while random spot-checks during data entry minimised transcription errors (Kalantari et al., 2024). For qualitative coding, two independent researchers analysed a subset of transcripts (20%). Inter-rater reliability, assessed using Cohen's kappa, was 0.81, indicating strong agreement (Li & Yu, 2022). Discrepancies were resolved through discussion before applying the coding framework to the full dataset.

Ethical Considerations

Ethical approval for this study was obtained from KIPPRA's Institutional Scientific Research and Ethics Committee. All participants provided informed consent before participation. Consent procedures included explaining the purpose of the study, assuring confidentiality, and obtaining written or verbal consent, depending on participants' context (Drolet et al., 2023). Respondents were informed of their right to withdraw at any stage without penalty.

Index Computation

The index was computed as follows:

For all dimensions considered, the scaled value for each of indicators were computed as

Scaled values for the indicators were computed using the formula

$$S_i = \frac{X_i - \text{Min}(x)}{\text{Max}(x) - \text{Min}(x)}$$

Where S_i = Scaled value for the indicator, and X_i = Data value of the indicator.

A composite indicator was computed to represent the dimension score using the formula:

$$\text{Weighted Index} = \frac{\sum W_i \cdot S_i}{\sum W_i}$$

Where W_i represents the unique weights associated with the scaled value S_i .

All indicators included in the study were assumed to be equally important in contributing to data governance in education.

The overall data governance index was computed using the formula.

$$\text{Weighted Index} = \frac{\sum W_i \cdot S_i}{\sum W_i}$$

Where W_i represents the unique weights associated with each dimension value, S_i

Pearson's Correlation Analysis was used to test the relationship between the data governance dimensions and the human capital index at the county level. Results

Status of Data Governance in Kenya

Kenya's overall Data Governance Index (DGI) score was **0.53**, reflecting a moderate level of maturity in national data governance. This marks an improvement from **0.39 in 2022** (Aaronson et al., 2022), indicating progress in institutionalising governance structures. Using the George Washington University (GWU) Framework for Public Sector Data Governance, Kenya was assessed across six pillars: strategic, regulatory, responsible, structural, participatory, and international.

- Kenya currently lacks a comprehensive National Data Strategy and Public Administration Data Strategy. A draft Artificial Intelligence Strategy (2025–2030) and the Data & Digital Transformation Strategy (2024–2028) indicate emerging efforts.
- Kenya has made notable progress, including the Data Protection Act (2019), the Access to Information Act (2016), and constitutional guarantees of freedom of information. Provisions on automated decision-making and data portability are also in place.
- Gaps remain significant, with no Data Charter, Public Sector Data Ethics Framework, or Trust Framework for Digital Identity Management. A Guidance Note on the Data Sharing Code (ODPC) is the only measure identified.
- Kenya has established the Office of the Data Protection Commissioner (ODPC) and launched the Kenya Open Data Initiative (KODI), but lacks a coordinating body for open data governance.
- Mechanisms for public consultation exist through the ODPC, but there is no formal system for government response to feedback or a multistakeholder advisory body.

- Kenya is a member of the Open Government Partnership and has adopted OECD AI Principles, but has not ratified Convention 108+, the Budapest Convention, or binding trade agreements on cross-border data flows.

Table 2 summarises the status of each indicator and the weighted scores used to compute Kenya's overall score of 0.53.

Table 2

Status of Data Governance in Kenya

Data Governance Attributes	Indicator	Status	Score	Weight	Weighted Score
Strategic (1/6)	National Data Strategy	No evidence	0	1/24	-
	Public Administration Data Strategy	No evidence	0	1/24	-
	AI Strategy	Draft (2025-2030)	1	1/24	0.04
	Strategy for Data in Emerging Digital Ecosystems	Data & Digital Transformation Strategy 2024–2028	1	1/24	0.04
Regulatory(1/6)	Personal Data Protection Law	Data Protection Act, 2019	1	1/30	0.03
	Open Data Law for the proactive release of government information	Kenya's Access to Information Act of 2016	1	1/30	0.03
	Freedom of Information Law	Article 35 of the Kenyan Constitution, which guarantees the right to access information.	1	1/30	0.03
	Right to be protected from Automated Decision-Making	Section 35 of the Kenya Data Protection Act of 2019	1	1/30	0.03
	Right to Data Portability	Data Protection Act, 2019	1	1/30	0.03
Responsible(1/6)	Data Charter	No	0	1/30	-
	Public Sector Data Ethics Framework	No	0	1/30	-
	Responsible AI Initiatives	No	0	1/30	-
	Trust Framework for Digital Identity Management	No- (Kenya Information Communications Act (1998, section 83E))	0	1/30	-
	Guidelines for Nongovernmental Data Sharing	Guidance Note on the Data Sharing Code (ODPC)	1	1/30	0.03
Structural(1/6)	Personal Data Protection Body	ODPC	1	1/24	0.04

	Open Data Portal	Kenya Open Data Initiative (KODI)	1	1/24	0.04
	Open Data Coordinating Body	No evidence	0	1/24	-
	Public Sector Data Governance Body	ODPC for personal data	1	1/24	0.04
Participatory(1/6)	Public Consultation on Data	Through the ODPC	1	1/18	0.06
	Government Response to Consultation	No evidence	0	1/18	-
	Multistakeholder Advisory Body	No evidence	0	1/18	-
International(1/6)	Convention 108+	Not a member	0	1/30	-
	Open Government Partnership	Yes- joined in 2011	1	1/30	0.03
	OECD AI Principles	Yes- adopted	1	1/30	0.03
	Binding Trade Agreements on Cross-Border Data Flows	No evidence	0	1/30	-
	Budapest Convention	No evidence	0	1/30	-
Total				1	0.53

Author's Computation based on the Digital Aaronson et al. (2022) Data Governance Framework

Data Governance Framework for Education

Kenya's education ecosystem generates data across learner records, teacher deployment, assessments, digital learning platforms, and institutional performance. A data governance framework defines decision rights, roles, processes, and controls for collecting, managing, sharing, and safeguarding this data. Core purposes include clarifying accountability, setting rules for data use, minimising risk, ensuring compliance, enhancing decision quality, improving communication among stakeholders, and increasing data value through accuracy, accessibility, and usability.

Five Pillars of Data Governance

Applying the UNDESA data governance framework (2025) to education data reveals five central pillars that underpin a well-functioning, trustworthy data ecosystem. These pillars ensure that education data is effectively managed, protected, and leveraged to improve learning outcomes and inform policy decisions. They include:

i. Roles and Responsibilities

The first pillar of education data governance in Kenya focuses on defining and institutionalizing roles and responsibilities across all actors involved in the data lifecycle. Key personnel include data stewards, data managers, data entry clerks, school administrators, and ICT officers at institutions, counties, and national bodies such as MoE, TSC, CUE, and KNEC. These professionals are the operational backbone of education data governance. Their responsibilities span the accurate collection, validation, storage, processing, and ethical use of learner and institutional data. Ensuring that these roles are clearly assigned and coordinated across systems such as NEMIS, KUCCPS, and institutional EMIS platforms is essential for building a trustworthy data ecosystem. Investment in digital literacy and data protection training for these

actors is critical. Without a competent and accountable workforce, even the best-designed data governance frameworks will fail to protect data subjects or enable data-driven decisions.

ii. Privacy Standards

The second pillar is the regulatory environment and privacy standards surrounding education data governance. The legal foundation is anchored in Kenya's Data Protection Act (2019), which applies across all educational institutions. Ensuring a lawful basis for collecting, processing, storing, and sharing personal education data is a prerequisite for ethical and compliant data practices. Key privacy safeguards include:

- (i) Ensuring informed consent for collection and use of student and staff data;
- (ii) Embedding privacy-by-design into EMIS and digital learning systems;
- (iii) Respecting the rights of data subjects, such as access, rectification, and erasure;
- (iv) Establishing clear data breach notification protocols;
- (v) Appointing trained Data Protection Officers at institutional levels;
- (vi) Conducting Data Protection Impact Assessments for new digital education systems.

These safeguards help ensure trust and protect learners whose data is increasingly being collected through biometric systems, learning apps, and centralized databases.

iii. Processes and Procedures

The third pillar outlines the standard operating procedures (SOPs) that govern how education data is collected, processed, secured, and shared. These SOPs establish the expected workflows, responsibilities, and checkpoints across the entire data lifecycle. Key areas include:

- i) Mapping of data flows between schools, county offices, and national databases;
- ii) Definitions of data quality standards (accuracy, timeliness, completeness, consistency);
- iii) Procedures for data validation, storage, and backup;
- iv) Mechanisms for data access control, version control, and audit trails;
- v) Continuous monitoring and improvement protocols to enhance data reliability.

Instituting these procedures ensures operational efficiency, enhances data quality, and strengthens institutional accountability.

iv. Policies and Guidelines

The fourth pillar involves creating and harmonizing policies and guidelines that govern how education data is managed, accessed, and shared. These policies establish the operational framework within which public institutions, private schools, TVETs, EdTech firms, and universities interact with education data. They also communicate the government's commitment to treating education data as a strategic national asset. Relevant policy tools include:

- (i) An internal data governance policy for ministries and institutions;

- (ii) An external (open) education data sharing policy, specifying conditions for third-party access;
- (iii) Data protection and information security guidelines, including minimum cybersecurity requirements for school IT systems;
- (iv) A data retention and disposal policy to manage the lifecycle of learner records.

While some of these elements exist in draft or fragmented form, a consolidated framework under the Ministry of Education is needed to standardise compliance and foster data use in planning, budgeting, and quality assurance.

v. Tools and Practices

This pillar focuses on the technical instruments that support the implementation of Pillar 4 policies. These tools enable institutions and data custodians to apply consistent data governance practices. They are grouped into two categories:

(a) Tools for Data Protection:

- i) Templates for anonymisation and pseudonymization of sensitive student data;
- ii) Guidance on data minimisation for surveys and assessments;
- iii) Standardised informed consent forms for parents and adult learners;
- iv) A framework for conducting institutional Data Protection Impact Assessments.

(b) Tools for Data Sharing and Access:

- i) A standardized code of conduct for data-sharing between ministries, research organisations, and EdTech firms;
- ii) Templates for data sharing agreements between institutions and third parties;
- iii) Tools for managing data inventories and access logs;
- iv) A data license registry outlining terms of use for external stakeholders accessing public education datasets.

These tools promote transparency, accountability, and structured data sharing while reducing legal and ethical risks.

Operationalising the Five Pillars

To evaluate the current state of data governance in Kenya's education sector, the five foundational pillars were assessed using six key operational dimensions. These dimensions provide a measurable lens for understanding the maturity, gaps, and strengths of existing governance structures. The six dimensions comprised a range of indicators to establish the status of each dimension from the perspectives of various key informants in the sector. Likert-type questions were used to assess the key informant's level of agreement with statements relating to the dimensions, based on documentary evidence (laws, policies, strategies, institutional structures).

- i) **Data Quality:** Evaluates the accuracy, completeness, consistency, and reliability of education datasets generated at school, county, and national levels. Poor data quality undermines the effectiveness of policy decisions and planning.
- ii) **Data Security:** Measures the presence and adequacy of technical and organizational safeguards to protect personal and institutional data from unauthorized access, loss, or breaches. This includes encryption, secure storage, and access controls aligned with the Data Protection Act.

- iii) **Data Accessibility:** Assesses how easily stakeholders can access relevant data when needed. It also considers issues of equity in access across urban and marginalized regions.
- iv) **Data Timeliness:** Examines the frequency, punctuality, and update cycles of data reporting. Timely data is essential for real-time decision-making, budget planning, and emergency response.
- v) **Data Stewardship:** Evaluates whether roles such as data stewards, custodians, and protection officers are designated and active at institutional and ministerial levels. This also includes clarity in accountability and oversight structures.
- vi) **Metadata Documentation and Organization:** Focuses on how well datasets are described, catalogued, and managed. Strong metadata practices enable better understanding, interoperability, and re-use of education data across systems.

Each of these assessment criteria aligns directly with the objectives of the five pillars and offers a structured approach to benchmark the current state of education data governance. Through this framework, it becomes possible to identify both systemic strengths and critical weaknesses.

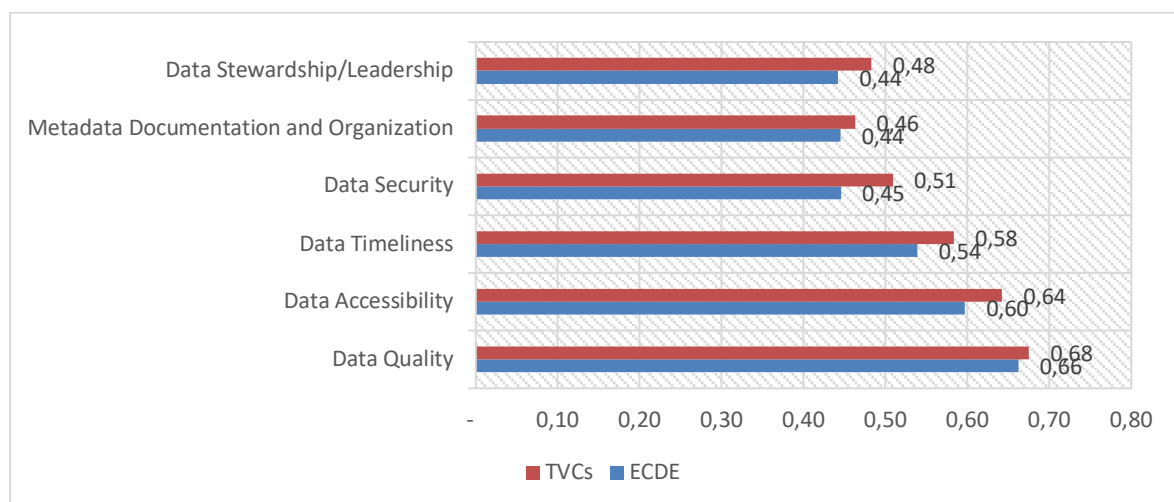
Data Governance Index (DGI) for Education Sector in Kenya

County level

The Data Governance Index scores indicate that county-managed Technical and Vocational Centers (TVCs) in Kenya outperform Early Childhood Development and Education (ECDE) programs across all six dimensions, though the margins are narrow (Figure 1). Data quality and accessibility are the strongest areas for both, largely due to counties adopting digital data collection tools such as NEMIS (National Education Management Information System), especially in TVCs. In counties like Makueni and Kakamega, structured reporting templates and digital records have improved data consistency and usability. Timeliness is moderately better in TVCs, where regular term-based assessments and county audits ensure more current data. In contrast, many ECDE programs rely on manual records, with limited update schedules, as seen in counties such as Turkana and Marsabit.

Figure 1

Data Governance Index at County Level Author's Computation based on the results of KIIs



Both sectors score poorly in metadata documentation, data security, and stewardship. Counties vary significantly in how they define and document datasets, leading to inconsistencies. For example, counties like Kiambu and Kisumu have begun standardising metadata through county data portals, while others lack

any formal classification systems. On security, most countries lack clear data protection policies or access controls, exposing both ECDE and TVC data to risks. Leadership in data governance also remains weak, with only a few counties, such as Nyeri and Elgeyo Marakwet, having designated data officers or units within the education departments. These findings underscore the need for a national framework guiding county-level practices, focused on strengthening metadata systems, securing infrastructure, and institutionalising data stewardship roles to ensure reliable, timely, and secure education data governance across Kenya.

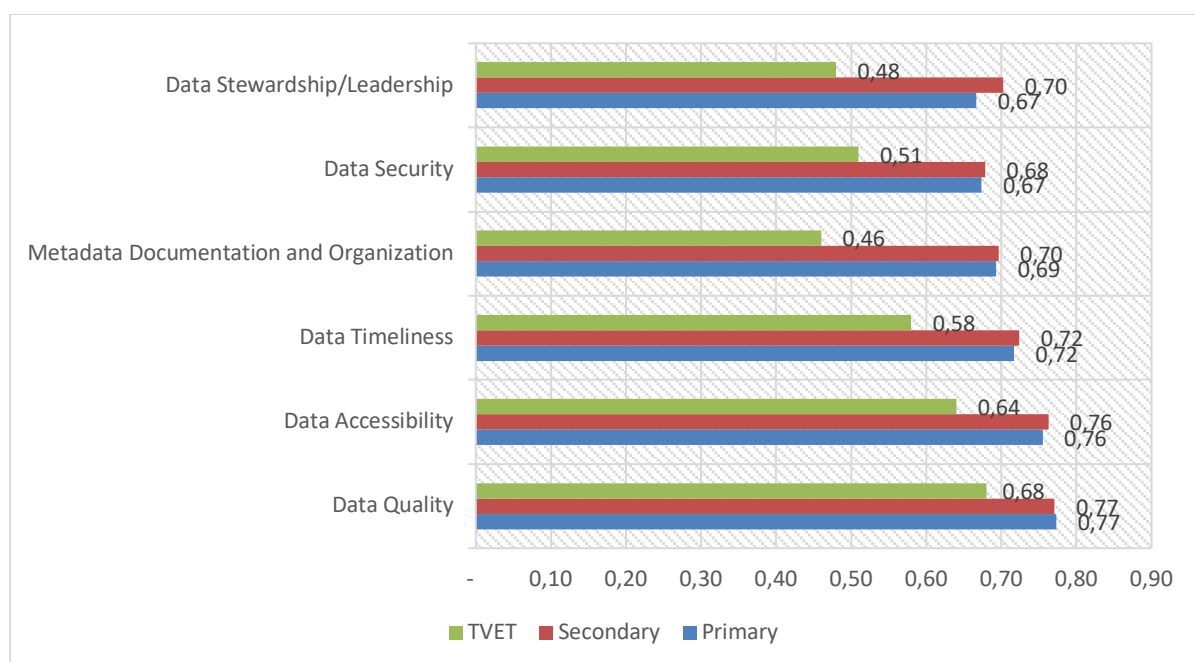
National Level

At the national level, the analysis revealed a clear disparity in data governance performance between national government-managed primary and secondary education institutions and TVET (Technical and Vocational Education and Training) institutions (Figure 2). Both primary and secondary sectors score highly and almost identically across all six dimensions, indicating that national-level systems, particularly those supported through NEMIS, have matured significantly. The consistent scores of 0.77 in data quality and 0.76 in data accessibility show that data in these sectors is not only reliable but also readily available for planning and accountability. The strong performance in data timeliness (0.72) reflects regular updates and efficient reporting cycles enforced by national standards.

In contrast, TVET institutions managed by national agencies score significantly lower in all dimensions, with particularly wide gaps in metadata documentation (0.46) and data stewardship (0.48). These gaps suggest a lack of standardised metadata practices and unclear governance roles, which are likely consequences of fragmented systems and the absence of a unified data management platform. While NEMIS provides comprehensive support for basic education, TVET relies more on sector-specific databases or manual tracking systems, which may not be uniformly implemented or maintained across institutions. This affects data reliability, consistency, and the capacity to support policy decisions or funding allocation efficiently.

Figure 2

Data Governance Index at National Level



Author's Computation based on the results of KIIs

The lower data security score (0.51) for TVET institutions is also notable and raises concerns about vulnerabilities in data protection, especially in institutions with limited ICT capacity. Similarly, weak data stewardship signals that leadership structures for managing education data in the TVET sector remain underdeveloped. These findings emphasize the need for national government intervention through the TVET Authority (TVETA) and Ministry of Education to invest in system integration, define clear governance roles, and develop a national data platform tailored to post-secondary training institutions. Aligning TVET data governance with the standards applied in basic education would significantly strengthen accountability, planning, and service delivery in Kenya's broader education system.

Data Quality and Human Development Index

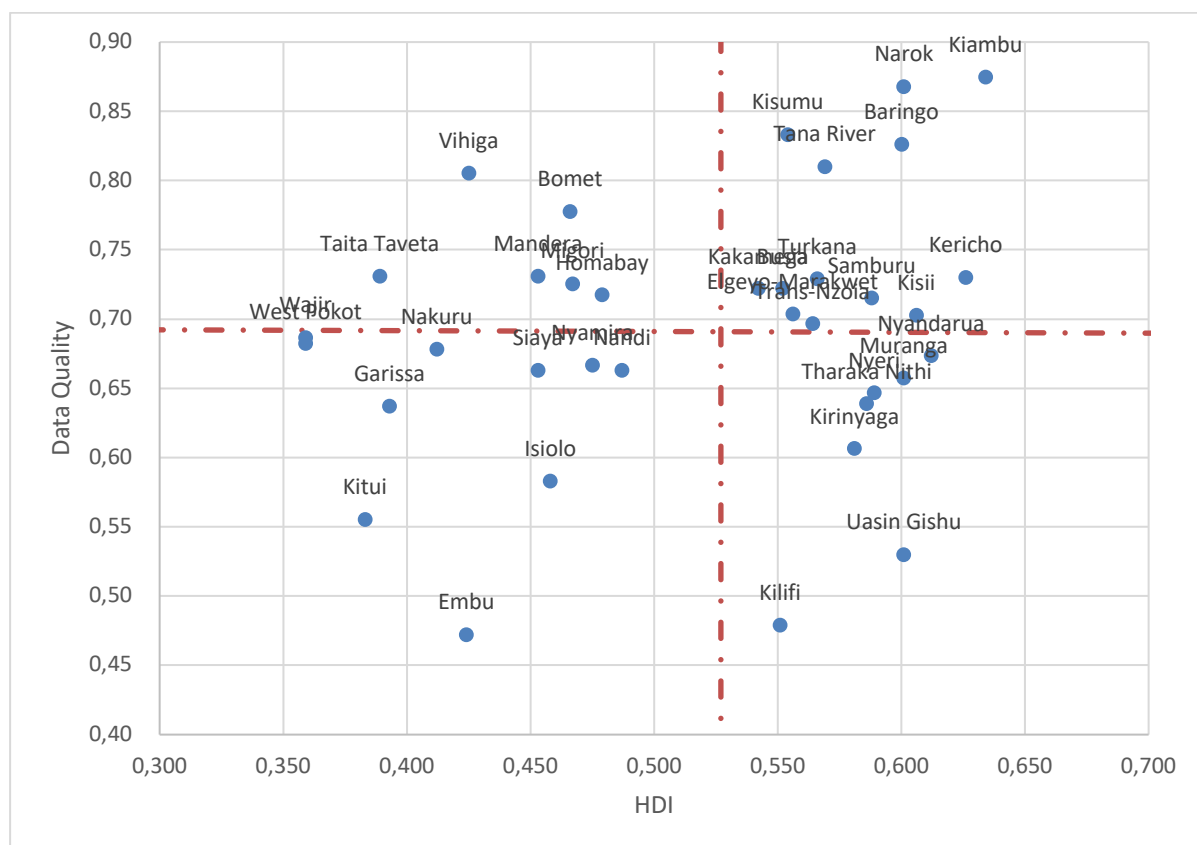
In this section, we focus on a component of the data governance index, namely data quality, and explore its correlation with HDI across counties. Data quality plays a vital role in advancing human development, as accurate, reliable, and timely data enables effective planning, resource allocation, and monitoring of services that directly impact HDI indicators such as education, health, and income. Figure 3 below shows a positive correlation between Data Quality dimension and the Human Development Index (HDI) across Kenyan counties. Counties with high HDI values, such as Kiambu (HDI 0.634, Data Quality 0.88) and Narok (HDI 0.601, Data Quality 0.87), also report very high data quality scores. These counties have invested in stronger institutional frameworks and technological tools that enable accurate, consistent, and actionable data. Such data quality enhances planning for healthcare, education, and infrastructure, contributing to better living standards, the very pillars of HDI. Conversely, counties like Embu (HDI 0.424, Data Quality 0.47) and Kilifi (HDI 0.551, Data Quality 0.48) show lower data quality, suggesting weak systems that may hinder evidence-based service delivery and, in turn, limit development outcomes.

At the national level, Kenya's average HDI stands at 0.527 with an average data quality score of 0.69, reflecting moderate alignment between development and data systems. Counties like Kericho, Bomet, and Tana River perform above this average in both indicators, indicating that sub-national governments can indeed surpass national standards when supported by effective leadership and resource mobilization. Interestingly, counties such as Vihiga (HDI 0.425, Data Quality 0.81) show that strong data systems may exist even where overall development is lower, possibly due to targeted support from NGOs or donor-funded programs enhancing data management without yet translating into overall human development gains.

The variations also highlight key national and county-level policy implications. Nationally, tools like NEMIS and government efforts to standardize data practices in education have benefited counties with better digital access and stronger governance. However, counties with low HDI and high data quality (e.g., Vihiga and Taita Taveta) reveal that data investments alone are not enough. They must be coupled with capacity-building, infrastructure, and equitable public investment to yield broader development results. Therefore, enhancing data quality remains essential not just for monitoring development but also as a driver of HDI improvements, especially when paired with responsive leadership and integrated development planning.

Figure 3

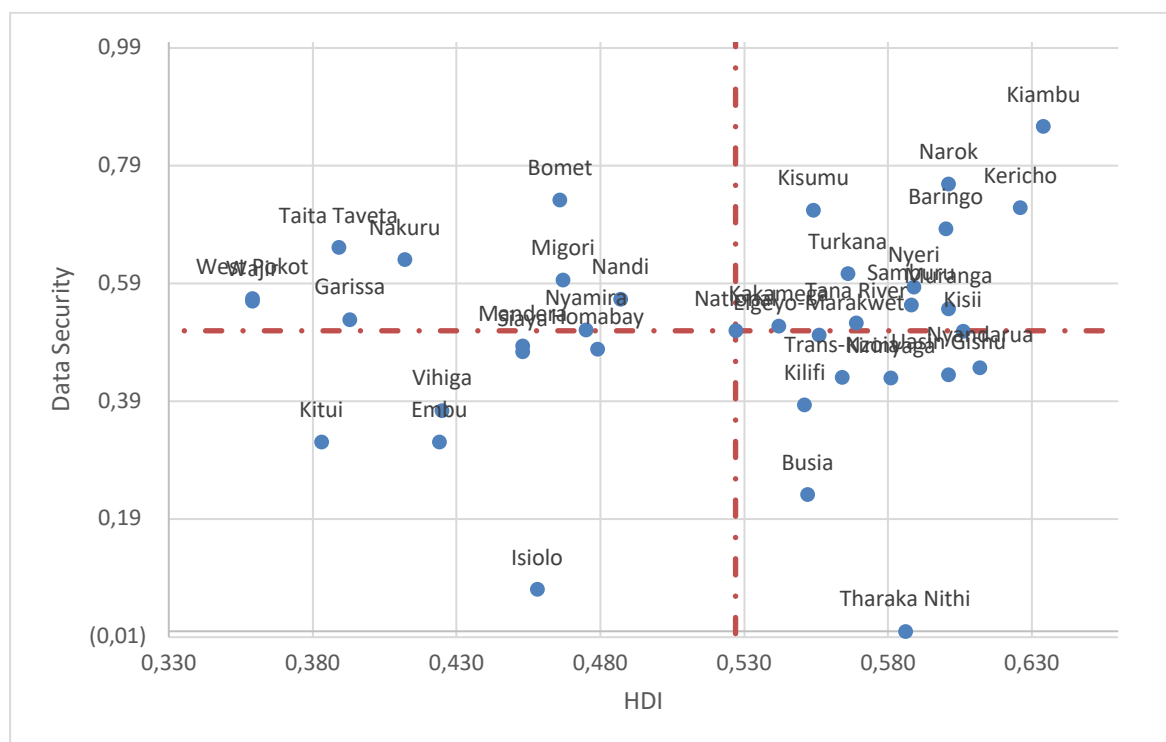
Data Quality and Human Development Index



Author's Computation based on the results of KIIs

Data Security and Human Development Index

In this section, we focus on the data security component of the data governance index and explore its correlation with HDI across countries. Data security is a critical enabler of human development, as it ensures the integrity, confidentiality, and trustworthiness of information used for planning and service delivery. A secure data environment strengthens public confidence, supports compliance with legal frameworks like Kenya's Data Protection Act (2012), and directly influences the quality of education, health, and governance services that shape the Human Development Index (HDI). The findings indicate significant disparities in data security practices across counties in Kenya, with most counties performing below or just at the national average score of 0.51. High-performing counties like Kiambu (0.86), Narok (0.76), and Kericho (0.72) exhibit strong adherence to key data security standards. These counties are likely to have implemented encryption, access control, routine security audits, and backup systems, as well as compliance with the Data Protection Act of 2012. In counties such as Kiambu, enhanced investments in ICT and governance capacity may explain their robust staff training programs, proactive incident response to breaches, and updated security protocols. Bomet (0.73) and Kisumu (0.71) similarly reflect counties where leadership prioritises secure digital systems, likely due to external donor support or integrated county ICT strategies.

Figure 4***Data Security and Human Development Index***

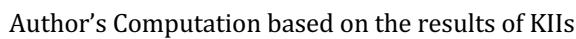
Author's Computation based on the results of KIIs

Conversely, counties such as Isiolo (0.07), Busia (0.23), and Embu (0.32) score extremely low, signalling significant vulnerabilities. These countries likely lack formalised data security structures, including encryption, regular training, or breach mitigation processes, exposing sensitive information to risk. In places like Isiolo, this may be due to limited infrastructure, low staffing capacity, or inadequate awareness of data protection laws. Even counties with mid-level HDI scores, such as Kilifi (HDI 0.551, Security 0.38) and Vihiga (HDI 0.425, Security 0.38), show weak enforcement of basic security measures. Overall, these findings highlight the urgent need for national coordination and county-level capacity-building to improve data security compliance, training, and resilience, especially in underserved or high-risk regions.

Data Accessibility and Human Development Index

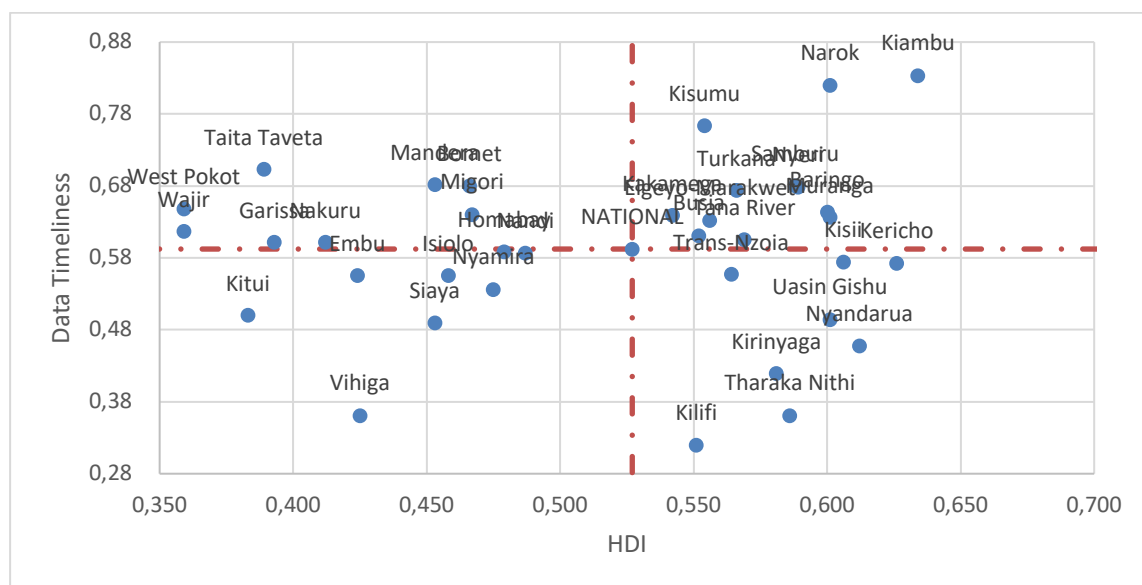
In this section, we focus on the data accessibility component of the data governance index and explore its correlation with HDI across counties. Data accessibility is a key driver of human development, as it enables timely and inclusive decision-making by ensuring that stakeholders can access, interpret, and utilise data effectively to improve education, health, and economic outcomes. The study revealed a generally positive relationship between data accessibility and HDI across Kenyan counties. High-performing counties such as Kiambu (HDI 0.634, Accessibility 0.93), Narok (HDI 0.601, Accessibility 0.87), and Kakamega (HDI 0.542, Accessibility 0.88) have invested in robust ICT infrastructure and user-friendly data systems, allowing both internal and external stakeholders to access updated information easily. These counties likely maintain digital repositories, provide regular training, and reduce legal or technical barriers to access, which enhances evidence-based planning. Kisumu (HDI 0.554, Accessibility 0.84) and Kericho (HDI 0.626, Accessibility 0.78) also show strong alignment, suggesting effective integration of digital tools and stakeholder support mechanisms that reinforce inclusive governance.

Data Accessibility and Human Development Index



Data Timeliness and Human Development Index

©Copyright 2025 by the author(s) This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 6***Data Timeliness and Human Development Index***

Author's Computation based on the results of KIIs

On the other end of the spectrum, counties like Kilifi (HDI 0.551, Timeliness 0.32), Tharaka Nithi (HDI 0.586, Timeliness 0.36), and Vihiga (HDI 0.425, Timeliness 0.36) reflect weaker systems likely characterised by staffing shortages, limited audit frequency, and low technology adoption, contributing to data delays. Counties such as Mandera (0.68), Bomet (0.68), and Taita Taveta (0.70), despite moderate HDI scores, have relatively strong data timeliness. This may be due to external collaboration with development partners or proactive county leadership introducing incentives for reporting and using mobile data collection tools. The national average of 0.59 suggests room for improvement across most counties. Enhancing mechanisms such as user feedback loops, real-time dashboards, and structured periodic audits will be essential to boost timeliness and, in turn, improve the planning and delivery of critical development services.

Data Stewardship/Leadership and Human Development Index

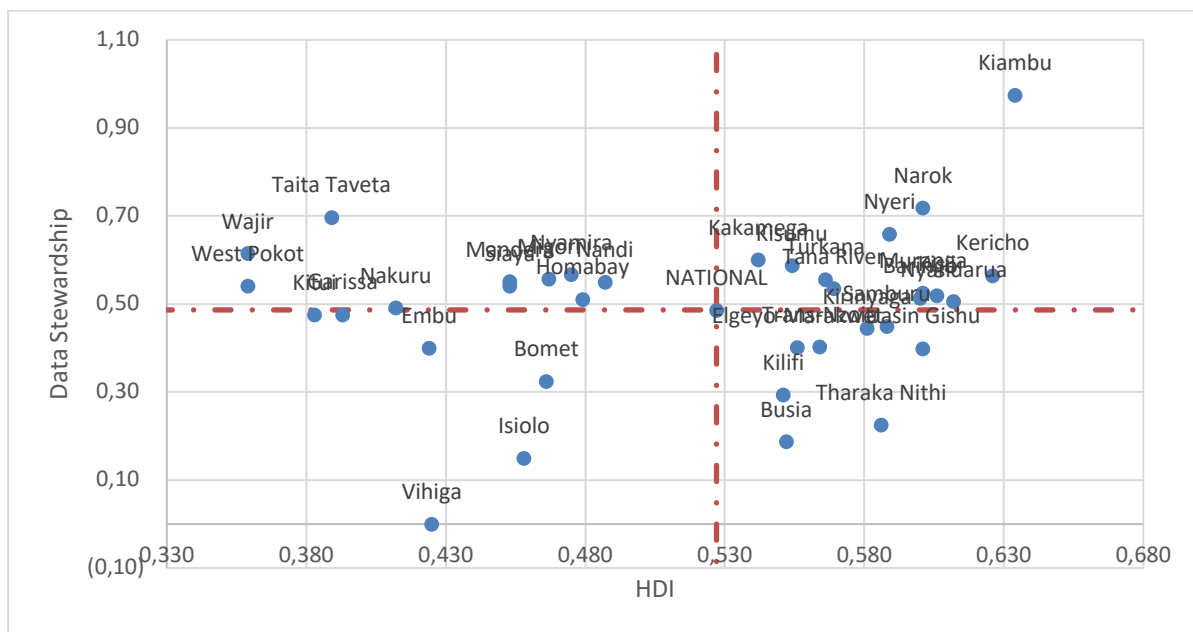
In this section, we focus on the data stewardship component of the data governance index and explore its correlation with HDI across counties. Strong data stewardship and leadership are essential for driving human development, as they ensure data is collected, managed, and used with transparency, accuracy, and accountability. Analysis of county-level data in Kenya shows a clear positive relationship between well-established data stewardship practices and higher Human Development Index (HDI) scores. Counties such as Kiambu (HDI 0.634, Stewardship 0.98), Narok (0.601, 0.72), and Nyeri (0.589, 0.66) have demonstrated strong governance by establishing clear roles, regular staff training, and functional education data management committees. These counties use reliable, timely data to inform policy and resource allocation, contributing to better health, education, and income outcomes.

In contrast, counties like Isiolo (HDI 0.458, Stewardship 0.15), Busia (0.552, 0.19), and Tharaka Nithi (0.586, 0.23) reveal significant weaknesses in data leadership. These gaps are often due to undefined responsibilities, limited funding, and a lack of operational frameworks, which hamper the effective use of data for development planning. While examples like Wajir (HDI 0.359, Stewardship 0.62) show that targeted support can improve data practices even in low-HDI contexts, the national average of 0.49 suggests a systemic need for investment in stewardship capacity. Strengthening data leadership through

policy frameworks, funding, and institutional roles is critical for improving development outcomes across all counties.

Figure 7

Data Stewardship/Leadership and Human Development Index



Author's Computation based on the results of KIIs

Metadata Documentation and Organisation and Human Development Index

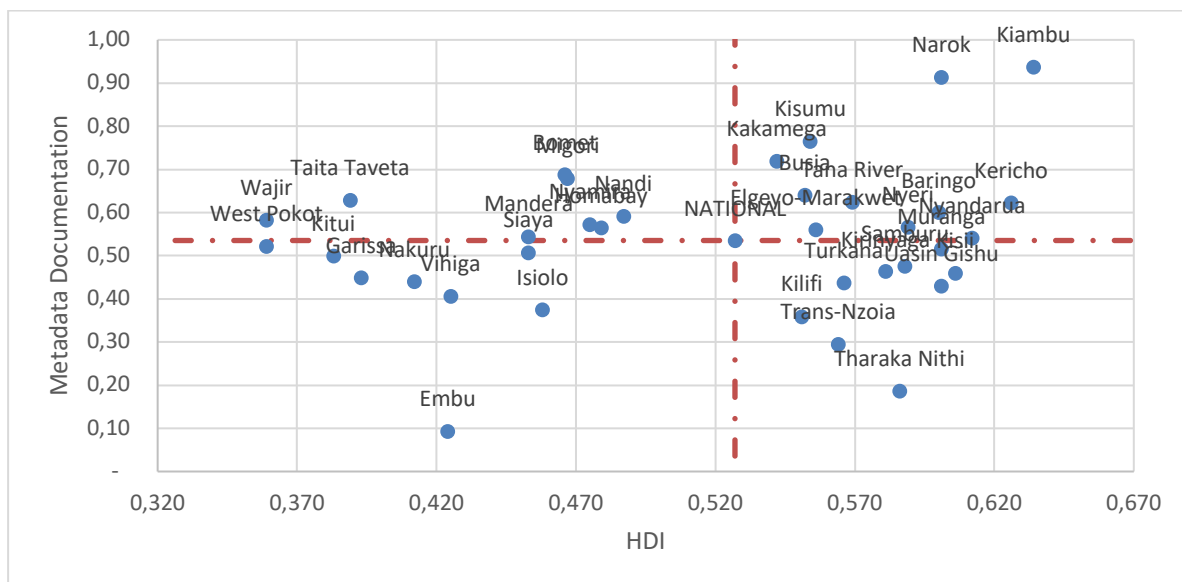
In this section, we focus on the metadata documentation component of the data governance index and explore its correlation with HDI across counties. Metadata documentation and organisation are foundational to high-quality data governance, as they enable clarity, consistency, and usability across all levels of the education system. Well-maintained metadata ensures users understand the meaning, structure, and source of data fields, supporting accurate analysis and policy decisions that directly impact development outcomes. In Kenya, counties with strong performance in this area, such as Kiambu (HDI 0.634, Metadata 0.94), Narok (HDI 0.601, 0.91), and Kisumu (HDI 0.554, 0.77), are likely benefiting from the use of standardised metadata frameworks, training of education staff, and established protocols for updating and using metadata. These counties may also employ mechanisms for user feedback to ensure that metadata practices evolve to meet the needs of various stakeholders. Their metadata coverage likely includes comprehensive documentation of data collection methods, sources, and definitions, contributing to more reliable education data and ultimately better human development outcomes.

Conversely, counties like Embu (HDI 0.424, Metadata 0.09), Tharaka Nithi (HDI 0.586, 0.19), and Trans-Nzoia (HDI 0.564, 0.30) reflect critical weaknesses in metadata systems. These counties may lack documented metadata standards, formal training for data users, or clear guidance for retrieving and interpreting data, leading to inconsistent, error-prone, or underutilised datasets. Nationally, Kenya's average metadata documentation score of 0.54 suggests that many counties operate without fully established systems, despite the Ministry of Education's efforts to promote structured data practices. While the National Education Management Information System (NEMIS) offers metadata support for primary and secondary schools, its reach and application in county-level TVET and ECDE data systems remain limited. Strengthening metadata organisation through the adoption of national metadata standards (such as ISO

11179), improving staff training, and enhancing system interoperability will be key to building a unified, actionable education data landscape that supports sustainable development across all countries.

Figure 8

Metadata Documentation and Organisation and Human Development Index



Author's Computation based on the results of KIIs

Data Governance Index and Human Development Index

In this section, we focus on the data governance index and explore its correlation with HDI across countries. The correlation analysis in Table 7 reveals a positive but modest relationship between HDI and all six dimensions of data governance, suggesting that improvements in data governance are associated with higher levels of human development across countries. The observed correlations between data governance dimensions and HDI align with existing literature emphasising the critical role of data systems in advancing human development outcomes, particularly in low- and middle-income countries. Scholars such as Kitchin (2014) and Jerven (2013) argue that high-quality, timely, and accessible data are foundational for effective public service delivery, policy formulation, and monitoring of socio-economic development. The relatively stronger correlation between data quality ($r = 0.24$) and HDI echoes findings by UNICEF (2017) and World Bank Group (2017), which underscore that robust Education Management Information Systems (EMIS) directly enhance the capacity of governments to target educational interventions and allocate resources equitably, thereby improving health, education, and income outcomes reflected in HDI scores.

Moreover, metadata documentation and accessibility, which also show notable associations with HDI, have been identified as key enablers of transparency, accountability, and evidence-based planning (World Bank, 2014; OECD, 2021). These dimensions facilitate interoperability and comparative analysis across regions, allowing sub-national governments to benchmark performance and replicate best practices. On the other hand, the weaker correlation observed for data timeliness ($r = 0.08$) is consistent with prior studies (e.g., Mo Ibrahim Foundation, 2019), which suggest that while timely data is crucial for real-time decision-making, its impact on long-term development outcomes may be mediated by institutional readiness and the presence of effective feedback loops.

Table 7*Pearson Correlations Between Data Governance Dimensions and HDI*

	1.	2.	3.	4.	5.	6.	7.	8.
1. HDI	1.00							
2. Data Quality	0.24	1.00						
3. Data Security	0.14	0.61	1.00					
4. Data Accessibility	0.19	0.75	0.80	1.00				
5. Data Timeliness	0.08	0.62	0.70	0.75	1.00			
6. Data Stewardship/Leadership	0.15	0.37	0.74	0.71	0.68	1.00		
7. Metadata Documentation and Organization	0.20	0.73	0.71	0.79	0.69	0.62	1.00	
8. Overall-Data Governance Index	0.19	0.75	0.90	0.92	0.85	0.83	0.88	1.00

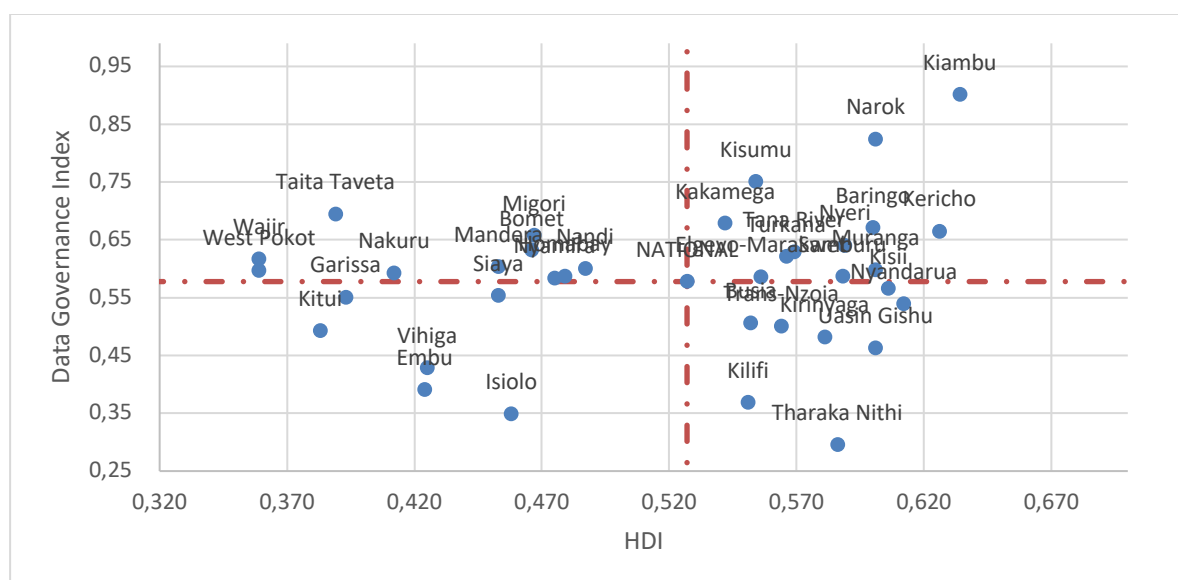
Authors' Computation

On the other hand, the strong internal correlations among the governance dimensions, as seen between data security and data accessibility ($r = 0.80$), reflect the integrated nature of data systems, where improvements in one area often reinforce improvements in others. This supports the integrated governance approach promoted by frameworks such as the OECD Data Governance Model (2021) and the African Union Digital Transformation Strategy (2020–2030), which advocate for systemic, coordinated data reforms to drive inclusive development. Thus, the empirical patterns observed in the Kenyan context are firmly rooted in and supported by global research on data governance and human development linkages.

Counties with stronger data governance frameworks tended to achieve higher levels of development. Counties such as Kiambu (HDI 0.634, DGI 0.90), Narok (HDI 0.601, DGI 0.82), and Kisumu (HDI 0.554, DGI 0.75) have some of the highest scores in both metrics. These counties have likely invested in integrated data systems, staff capacity-building, robust stewardship mechanisms, and routine data audits. As a result, their ability to generate accurate, timely, and actionable data contributes directly to improved planning, transparency, and service delivery in education, health, and other human development sectors.

Figure 9

Data Governance Index and Human Development Index



Author's Computation based on the results of KIIs

Conversely, counties like Tharaka Nithi (HDI 0.586, DGI 0.29), Isiolo (HDI 0.458, DGI 0.35), and Kilifi (HDI 0.551, DGI 0.37) illustrate how weak data governance structures may constrain development efforts. These counties may lack standardised metadata, face irregular data reporting, or operate with minimal oversight and accountability in education data systems. The national average DGI of 0.58 indicates moderate performance overall, but also highlights the need for targeted policy interventions to bridge gaps in lower-performing counties. Notably, counties like Taita Taveta (HDI 0.389, DGI 0.69) and West Pokot (HDI 0.359, DGI 0.60) show that even in regions with lower HDI, strong governance practices can emerge, often due to focused donor support or local leadership reforms. This divergence between relatively strong data governance and lower HDI scores in certain countries invites closer examination. For instance, counties like Taita Taveta (HDI 0.389, DGI 0.69) and West Pokot (HDI 0.359, DGI 0.60) demonstrate that robust data governance practices can exist even in regions with lower overall development outcomes. This may be attributable to targeted donor interventions, pilot initiatives, or progressive county-level leadership that prioritise digital infrastructure and data reforms regardless of broader socio-economic challenges. However, the absence of a strong correlation in these cases suggests that improved data systems alone may not be sufficient to immediately shift HDI metrics, which are influenced by long-term structural factors such as poverty, infrastructure, and access to health care. It would therefore be valuable to track the HDI trajectories of such countries over time to determine whether sustained data governance efforts eventually lead to corresponding gains in human development. This could provide critical insights into lag effects and help refine policy timelines and expectations for reform impact. Strengthening county-level data governance across all dimensions will be essential for achieving equitable and sustainable development in Kenya.

Discussion

Status of Data Governance in Kenya

Effective data governance is a critical enabler of digital transformation, ensuring accountability and fostering inclusive development (Paranahewage et al., 2025; Nookala, 2024; Onoja et al., 2024). In an

increasingly data-driven world, a country's ability to manage, protect, and leverage its data assets determines the effectiveness of its policies, the trust of its citizens, and the competitiveness of its economy. While global frameworks offer guiding principles, each country faces unique legal, institutional, technological, and socio-political contexts that shape how data governance should be designed and implemented. Tailoring governance systems to national realities ensures that data policies are not only compliant with international norms but also responsive to local needs (Fischer, 2023). For developing countries like Kenya, strengthening data governance is especially vital to enhance public service delivery, monitor development outcomes, safeguard privacy, and attract data-driven investment (Pemmasani et al., 2022).

Kenya's data governance landscape demonstrates measurable progress, particularly in regulatory and structural dimensions. The enactment of the Data Protection Act (2019) and the establishment of the ODPC provide a strong legal and institutional foundation for protecting personal data and safeguarding citizen rights. Similarly, initiatives like KODI promote transparency and public access to government data. Despite these advances, fragmentation remains a persistent challenge. The absence of a centralised coordinating body hampers harmonisation across ministries and agencies, limiting consistency and efficiency. This suggests that Kenya's progress has mainly been regulatory rather than systemic.

The strategic and responsible pillars are underdeveloped, with no national-level strategies or ethics frameworks in place. Compared with leading countries such as Estonia or Canada, which embed trust, ethics, and multi-stakeholder engagement into governance, Kenya's system remains incomplete. On the participatory front, the ODPC provides avenues for consultation but lacks mechanisms to respond to feedback or engage stakeholders systematically. This shortfall highlights the need for inclusivity in governance structures, as participatory oversight is a critical feature of international best practice. In terms of international commitments, Kenya has aligned with the OECD AI Principles and joined the Open Government Partnership. However, its absence from binding agreements like Convention 108+ and the Budapest Convention limits its integration into global digital governance frameworks. This gap underscores the importance of aligning national frameworks with international standards to enhance cross-border trust and cooperation (Wong Villanueva et al., 2022). Overall, Kenya's upward trajectory from 0.39 in 2022 to 0.53 in 2025 suggests growing institutionalisation of laws and policies. While still trailing global leaders, the country is positioning itself as a regional example of emerging governance capacity. Future reforms should focus on strategic planning, ethical oversight, participatory governance, and international integration to build a comprehensive, human-centred system.

Data Governance Framework for Education

The DGI provides the first comprehensive measurement of education data governance in Kenya, offering essential insights into how data governance structures intersect with human development. The results confirm uneven performance across the six operational dimensions and highlight both national progress and persistent county-level disparities. Below, the findings are situated within international benchmarks, interpreted in terms of statistical and practical significance, and reflected upon in light of methodological limitations and unexpected patterns.

The correlation analysis revealed positive but modest associations between DGI dimensions and HDI, with coefficients ranging from $r = 0.08$ (timeliness) to $r = 0.24$ (data quality). Confidence intervals at the 95 per cent level (e.g., data quality-HDI, 0.09 to 0.39; metadata-HDI, 0.06 to 0.34) confirm that these relationships are statistically significant, albeit not large. These results align with prior studies that emphasise the enabling but indirect role of governance structures in shaping development outcomes. For instance, Samoilovych et al. (2022) established a positive relationship between human capital development and the expansion of the digital economy, while Jallow and Sanner (2023) reported that stronger education

management information systems improve equity in resource allocation and service delivery. Similarly, Jerven (2013) and Kitchin (2014) argue that the reliability and accessibility of data provide essential conditions for effective governance, even if their immediate impact on development indicators remains modest. Statistically, the findings suggest that data governance contributes incrementally rather than dramatically to human development. Practically, however, the implications are significant: counties with stronger governance systems consistently outperformed peers in planning efficiency, accountability, and service delivery. Even small gains in DGI dimensions can enhance education resource allocation, improve monitoring, and strengthen citizen trust, which, over time, are critical for long-term human capital development.

The cross-sectional design constrains the ability to draw causal inferences. Correlations observed between DGI and HDI do not establish directionality. Stronger human development could facilitate better governance through increased capacity, or conversely, improved governance could drive better development outcomes (Poquet & De Laat, 2021; Mou et al., 2022). Longitudinal data are needed to capture lag effects, in which improvements in governance structures today may influence HDI scores years later. The cross-sectional nature of this study, therefore, limits causal claims and should be interpreted as evidence of associations rather than direct effects.

A notable and unexpected result is the presence of counties with high DGI but low HDI, including Taita Taveta (DGI 0.69, HDI 0.389) and West Pokot (DGI 0.60, HDI 0.359). This divergence highlights the complexity of the governance–development relationship. Possible explanations include donor-funded programs that prioritised governance reforms without equivalent investments in broader socio-economic infrastructure, or progressive county leadership introducing governance innovations that have not yet translated into measurable development outcomes. These cases underscore the need to track trajectories over time and to integrate data governance reforms with broader investments in infrastructure, poverty alleviation, and service delivery.

Three key lessons emerge. First, counties with strong DGI scores but stagnant HDI should be targeted with integrated interventions that pair governance reforms with direct investments in social services. Second, low-performing counties in both metrics, such as Isiolo and Kilifi, require urgent foundational support in ICT infrastructure, training, and legal enforcement. Third, benchmarking against global leaders suggests that prioritising metadata documentation, stewardship, and security would yield the highest returns for Kenya’s progression toward international norms. Institutionalising data officers at the county level and adopting ISO-aligned metadata standards are concrete steps forward.

Conclusion

The analysis of Kenya’s Data Governance Index reveals a strong correlation between robust data governance practices and improved Human Development Index (HDI) outcomes. Counties with well-structured data systems characterised by high data quality, accessibility, timeliness, security, metadata documentation, and stewardship demonstrate stronger planning capabilities, more efficient service delivery, and better outcomes in health, education, and economic sectors. High-performing counties like Kiambu, Narok, and Kisumu exemplify how institutionalised data leadership, regular audits, digital tools, and user engagement can collectively drive meaningful development. Conversely, counties with weaker data governance often struggle with data inconsistencies, limited capacity, and poor coordination, which hamper evidence-based decision-making and equitable progress. Theoretically, the findings extend data governance literature by positioning education data systems as mediators between institutional structures and human capital formation. In doing so, the study builds a bridge between Human Capital Theory and digital governance, showing how investments in data quality, accessibility, and stewardship enable education systems to convert resources into developmental gains more effectively.

To achieve sustained national development, Kenya must prioritise investment in comprehensive data governance reforms across all counties. This includes adopting national standards, strengthening institutional roles, ensuring consistent training and funding for data activities, and enforcing compliance with existing laws such as the Data Protection Act (2012). A harmonised, inclusive, and well-resourced data governance ecosystem will not only improve transparency and accountability but also empower counties to better respond to citizens' needs. Strengthening these foundations is essential for closing development gaps and ensuring that all counties, regardless of starting point, have the tools and systems necessary to contribute to Kenya's Vision 2030 and the Sustainable Development Goals. The results also carry relevance beyond Kenya. Many Sub-Saharan African countries face similar challenges of fragmented data systems, weak metadata documentation, and limited stewardship capacity. The Kenyan case provides a replicable framework for benchmarking data governance. It offers lessons on how targeted reforms in metadata standards, role definition, and security protocols can strengthen accountability and planning in resource-constrained contexts.

Specific recommendations follow for different stakeholders. For national policymakers, adopting national metadata standards, institutionalising data stewardship roles, and harmonising TVET and ECDE platforms are priorities. For county governments, investments in ICT infrastructure, staff training, and open data portals can reduce disparities. For development partners and donors, support should focus on building long-term institutional capacity rather than short-term system upgrades. Together, these actions can enhance transparency, equity, and trust in education data governance across the region.

A summary of data governance issues and recommendations is provided in Table 2.

Recommendations

To strengthen Kenya's education data ecosystem, it is essential to focus on the most critical challenges that undermine effective data governance across counties and institutions. These include inconsistent data quality, weak metadata documentation, limited accessibility, gaps in data security, and the absence of strong stewardship structures. Addressing these issues is vital to ensure reliable evidence for decision-making, build public trust, and support equitable delivery service in the education sector. Based on the analysis, the refined recommendations have been consolidated into three priority areas: strengthening data quality and metadata systems, enhancing accessibility and security, and institutionalising stewardship and leadership. These proposals aim to improve data reliability, safeguard privacy, expand equitable access, and establish sustainable leadership structures, while clearly assigning roles to responsible actors at both national and county levels. This prioritisation supports the development of a cohesive and responsive education data governance framework aligned with Kenya's national development goals and global best practices.

Table 8

Data Governance Issues and Recommendations

Priority Area	Rationale	Key Actions	Lead Actors
1. Strengthen Data Quality & Metadata Systems	Strongest correlation with HDI; foundation for reliability, comparability, and actionable education data.	- Implement mandatory validation protocols and automate Data collection. - Adopt national metadata standards and create centralized repositories.	MoE Quality Assurance, KNBS, County Education Directors, EMIS Officers

		• Train staff in data accuracy, integrity, and metadata documentation.	
2. Enhance Accessibility & Security	Accessibility ensures data use for planning and equity; security builds public trust and compliance with the Data Protection Act.	• Develop open-access platforms and harmonized policies across counties. • Enforce encryption, access controls. • Provide regular training on data access and cybersecurity.	MoE ICT Directorate, ODPC, County ICT Teams, Civil Society & Researchers
3. Institutionalize Stewardship & Leadership	Ensures sustainability and accountability of reforms; dedicated roles and budgets prevent fragmentation.	• Establish stewardship roles and cross-sector committees. • Allocate dedicated budgets for data governance at county and national levels. • Promote accountability and transparency through audits and reporting.	MoE EMIS Directorate, County Directors of Education, County Executive Committees

Author's Compilation

Limitations

This study uses a cross-sectional design, which allows identification of associations between the Data Governance Index (DGI) and the Human Development Index (HDI) but does not capture long-term causal dynamics (Savitz & Wellenius, 2023; Imai et al., 2021). Some indicators, particularly for stewardship and metadata documentation, relied on qualitative assessments, which may reduce precision but still provide useful benchmarks for comparison (Ahmed, 2024). Finally, the differences in institutional maturity across sub-sectors also mean that national systems such as NEMIS appear more advanced than ECDE or TVET platforms. This disparity reflects Kenya's reform trajectory rather than flaws in measurement.

Suggestions for future research

This study provides a foundation for understanding education data governance in Kenya, yet future research could expand and deepen this work in several directions. First, longitudinal designs are needed to examine how improvements in data governance translate into human development outcomes over time. Such designs would help clarify lag effects that cross-sectional analyses cannot capture, particularly in countries where governance reforms are evident, but HDI gains remain limited. Second, comparative studies across Sub-Saharan Africa could situate Kenya's experience within broader regional patterns, thereby identifying shared challenges and scalable best practices. Third, integrating advanced quantitative approaches, such as multilevel modelling or structural equation modelling, with qualitative case studies would generate richer insights into the institutional, political, and technological drivers of data governance outcomes. Fourth, future studies could focus on underexplored sub-sectors, especially TVET and ECDE, where governance structures remain less developed, and data systems are fragmented. Finally, methodological innovation is needed to create standardised tools for measuring complex dimensions such as stewardship, metadata documentation, and participatory governance. Developing such instruments would enhance reliability, allow cross-country benchmarking, and support policymakers in tracking progress toward international norms. Collectively, these research avenues can strengthen the evidence base for building inclusive and accountable education data ecosystems.

Acknowledgements

The authors gratefully acknowledge the African Economic Research Consortium (AERC) for funding this study and providing valuable support throughout the research process. We also extend our sincere appreciation to the Kenya Institute for Public Policy Research and Analysis (KIPPRA), where the study was conducted. The institutional environment, guidance, and resources offered by KIPPRA were instrumental in shaping the completion of this work.

Conflict of Interest

None

Funding

This study was funded by the African Economic Research Consortium (AERC) under a research grant provided to support the identification of institutional structures for data policy and governance frameworks that can be replicated across Sub-Saharan African countries. The grant facilitated fieldwork, data collection, and analysis. The Kenya Institute for Public Policy Research and Analysis (KIPPRA) provided the institutional environment where the study was conducted.

References

- Aaronson, S., Struett, T., & Zable, A. (2022). *DataGovHub paradigm for a comprehensive approach to data governance: Year 1 report*. The Digital Trade & Data Governance Hub, George Washington University; IDEA 2030 Council—Digital Planet, Fletcher School at Tufts University. <https://digitalplanet.tufts.edu/wp-content/uploads/2022/09/DataGovHub-Paradigm-for-a-Comprehensive-Approach-to-Data-Governance-Y1.pdf>
- African Union. (2020). *Digital Transformation Strategy for Africa (2020–2030)*. African Union Commission. <https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf>
- Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health*, 2, 100051. <https://doi.org/10.1016/j.glmedi.2024.100051>
- Barrera-Osorio, F., & Linden, L. (2009). *The use and misuse of computers in education: Evidence from a randomized experiment in Colombia*. World Bank Group. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/346301468022433230>
- Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis with special reference to education*. University of Chicago Press. <https://www.nber.org/system/files/chapters/c3730/c3730.pdf>
- Briggs, A., & Cafaro, F. (2021). End-user needs of fragmented databases in higher education data analysis and decision making. *Informatics*, 8(3), 42. <https://doi.org/10.3390/informatics8030042>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679. <https://doi.org/10.1257/aer.104.9.2633>
- Drolet, M. J., Rose-Derouin, E., Leblanc, J. C., Ruest, M., & Williams-Jones, B. (2023). Ethical issues in research: Perceptions of researchers, research ethics board members and research ethics experts. *Journal of Academic Ethics*, 21(2), 269–292. <https://doi.org/10.1007/s10805-022-09455-3>

- Fischer, A. (2023). Data sovereignty and e-governance: The legal implications of national laws on digital government systems. *Legal Studies in Digital Age*, 2(4), 1–12. <https://www.jlsda.com/index.php/lstda/article/view/39/41>
- Imai, K., Kim, I. S., & Wang, E. H. (2023). Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 67(3), 587–605. <https://doi.org/10.1111/ajps.12685>
- Jallow, S. A., & Sanner, T. A. (2023). The role of EMIS in improving equitable service delivery in education: A case study. In *After Latour: Globalisation, inequity and climate change* (Vol. 696, p. 140). Springer. https://doi.org/10.1007/978-3-031-50154-8_5
- Jerven, M. (2013). *Poor numbers: How we are misled by African development statistics and what to do about it*. Cornell University Press. <https://doi.org/10.7591/9780801467615>
- Jim, C. K., & Chang, H. C. (2018). The current state of data governance in higher education. *Proceedings of the Association for Information Science and Technology*, 55(1), 198–206. <https://doi.org/10.1002/pr2.2018.14505501022>
- Kalantari, R., Pakravan-Charvadeh, M. R., & Rahimian, M. (2024). Multi-level factors influencing climate migration willingness among small-scale farmers. *Frontiers in Environmental Science*, 12, 1434708. <https://doi.org/10.3389/fenvs.2024.1434708>
- Kenya National Bureau of Statistics. (2024). *Statistical abstracts*. <https://www.knbs.or.ke/wp-content/uploads/2024/12/2024-Statistical-Abstract1.pdf>
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. SAGE Publications. <https://doi.org/10.4135/9781473909472>
- Li, M., & Yu, T. (2022). Methodological issues in evaluating agreement between two detection methods by Cohen's kappa analysis. *Parasites & Vectors*, 15(1), 270. <https://doi.org/10.1186/s13071-022-05402-8>
- Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, 51(3), 485–504. <https://doi.org/10.1002/ejsp.2752>
- Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α , McDonald's ω t and the greatest lower bound. *Social Sciences & Humanities Open*, 7(1), 100368. <https://doi.org/10.1016/j.ssaho.2022.100368>
- Materu, P. (2007). *Higher education quality assurance in Sub-Saharan Africa: Status, challenges, opportunities, and promising practices*. World Bank. <https://openknowledge.worldbank.org/server/api/core/bitstreams/e9904501-5711-52ae-86b7-b860bb730adb/content>
- McCarthy, A. M., Maor, D., McConney, A., & Cavanaugh, C. (2023). Digital transformation in education: Critical components for leaders of system change. *Social Sciences & Humanities Open*, 8(1), 100479. <https://doi.org/10.1016/j.ssaho.2023.100479>
- Mo Ibrahim Foundation. (2019). *Assessing Africa's data landscape*. https://mo.ibrahim.foundation/sites/default/files/2019-11/African_Governance_Report_2019_EN.pdf

- Mou, A. J., Hossain, M. S., & Siddiqui, N. A. (2022). Digital transformation in marketing: Evaluating the impact of web analytics and SEO on SME growth. *American Journal of Interdisciplinary Studies*, 3(4), 61–90. <https://doi.org/10.63125/8t10v729>
- Nookala, G. (2024). Adaptive data governance frameworks for data-driven digital transformations. *Journal of Computational Innovation*, 4(1). <https://researchworkx.com/index.php/jci/article/view/16/16>
- Odhiambo, N. M. (2024). Education and economic growth in Sub-Saharan African countries: Does governance quality matter? *Research in Globalization*, 8, 100227. <https://doi.org/10.1016/j.resglo.2024.100227>
- OECD. (2021). *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*. OECD Publishing. <https://doi.org/10.1787/276aaca8-en>
- OECD. (2022). *Going digital to advance data governance for growth and well-being: Data governance for data-driven innovation*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/12/going-digital-to-advance-data-governance-for-growth-and-well-being_246d8cab/e3d783b0-en.pdf
- Ojiemhenkele, A. E., & Ofangbonmu, U. (2024). Exploring the panaceas to the consequences of inadequate data to the planning of education in Nigeria. *Journal of Theoretical and Empirical Studies in Education*, 9(1), 246–265. <https://journals.unizik.edu.ng/jtese/article/view/5050/4179>
- Onoja, J. P., Hamza, O., Collins, A., Chibunna, U. B., Eweja, A., & Daraojimba, A. I. (2021). Digital transformation and data governance: Strategies for regulatory compliance and secure AI-driven business operations. *Journal of Frontiers in Multidisciplinary Research*, 2(1), 43–55. <https://doi.org/10.54660/IJFMR.2021.2.1.43-55>
- Osei-Kusi, F., Wu, C. S., & Tetteh, S. (2024). Unlocking economic growth and sustainable development in Sub-Saharan Africa. *Heliyon*, 10(12). <https://doi.org/10.1016/j.heliyon.2024.e33252>
- Paranahewage, P., & Lakdinu, K. (2025). *Data governance's role in digital transformation: A systematic literature review*. <https://www.diva-portal.org/smash/get/diva2:1967803/FULLTEXT01>
- Pemmasani, P. K., & Abd Nasaruddin, M. A. (2022). Strengthening public sector data governance: Risk management strategies for government organizations. *International Journal of Modern Computing*, 5(1), 108–118. <https://yuktabpublisher.com/index.php/IJMC/article/view/250/265>
- Poquet, O., & De Laat, M. (2021). Developing capabilities: Lifelong learning in the age of AI. *British Journal of Educational Technology*, 52(4), 1695–1708. <https://doi.org/10.1111/bjet.13123>
- Raudonyte, I., & Foimapafisi, T. (2022). *Using learning assessment data for educational planning in Sub-Saharan Africa: A comparative analysis*. UNESCO International Institute for Educational Planning. <https://coilink.org/20.500.12592/gq8z5g>
- Samoilovych, A., Popelo, O., Kychko, I., Samoilovych, O., & Olyfirenko, I. (2022). Management of human capital development in the era of the digital economy. *Journal of Intelligent Management Decisions*, 1(1), 56–66. <https://doi.org/10.56578/jimd010107>
- Savitz, D. A., & Wellenius, G. A. (2023). Can cross-sectional studies contribute to causal inference? It depends. *American Journal of Epidemiology*, 192(4), 514–516. <https://doi.org/10.1093/aje/kwac037>

- Schlunegger, M. C., Zumstein-Shaha, M., & Palm, R. (2024). Methodologic and data-analysis triangulation in case studies: A scoping review. *Western Journal of Nursing Research*, 46(8), 611–622. <https://doi.org/10.1177/01939459241263011>
- Schultz, T. W. (1961). Investment in human capital. *American Economic Review*, 51(1), 1–17. <https://la.utexas.edu/users/hcleaver/330T/350kPEESchultzInvestmentHumanCapital.pdf>
- Tchamy, V. S., Asongu, S. A., & Odhiambo, N. M. (2019). The role of ICT in modulating the effect of education and lifelong learning on income inequality and economic growth in Africa. *African Development Review*, 31(3), 261–274. <https://doi.org/10.1111/1467-8268.12388>
- UNICEF. (2017). *Data for children strategic framework*. United Nations Children's Fund. <https://data.unicef.org/wp-content/uploads/2017/04/Data-for-Children-Strategic-Framework-UNICEF.pdf>
- United Nations Department of Economic and Social Affairs. (2025). *Capacity development and consultation workshop on public service management and data governance: Day one presentation*. Division for Public Institutions and Digital Government. <https://publicadministration.desa.un.org/sites/default/files/list-of-files/2025/1.1.2%29%20JNT%20Day%20One%20Presentation%20UNDESA%20Workshop.pdf>
- Wamutoro, M., Kessio, D. K., & Wambua, B. K. (2022). Effectiveness of EMIS for student information management in public secondary schools in Uasin Gishu County, Kenya. *Reviewed Journal International of Business Management*, 3(1), 122–133. <https://doi.org/10.36348/jaep.2022.v06i11.002>
- Wang, Y., Hong, D., & Huang, J. (2023). A diffusion of innovation perspective for digital transformation in education. *Procedia Computer Science*, 225, 2439–2448. <https://doi.org/10.1016/j.procs.2023.10.235>
- Watson, C. (2019). Context and implications document for: From accountability to digital data: The rise and rise of educational governance. *Review of Education*, 7(2), 428–429. <https://doi.org/10.1002/rev3.3131>
- Wong Villanueva, J. L., Kidokoro, T., & Seta, F. (2022). Cross-border integration, cooperation and governance: A systems approach for evaluating “good” governance in cross-border regions. *Journal of Borderlands Studies*, 37(5), 1047–1070. <https://doi.org/10.1080/08865655.2020.1855227>
- World Bank Group. (2017). *Data for learning: Building a smart education data ecosystem*. World Bank. <https://openknowledge.worldbank.org/bitstreams/7661f023-72c7-5ece-be00-75921a193562/download>
- World Bank. (2014). *Open Data for Resilience Initiative: Field guide*. The World Bank. <https://openknowledge.worldbank.org/server/api/core/bitstreams/81009d98-8871-54be-98c1-be605001a7b6/content>
- World Bank. (2018). *World Development Report 2018: Learning to realize education's promise*. World Bank. <https://www.worldbank.org/en/publication/wdr2018>
- World Economic Forum. (2023). *How Africa's youth will drive global growth*. <https://www.weforum.org/stories/2023/08/africa-youth-global-growth-digital-economy/>