
Challenges in New Data Sources and Opportunities in Kenya

By

Ms. Leah Wambugu
National Statistical System
Kenya National Bureau of Statistics (KNBS)
lwambugu@knbs.or.ke

Regional Conference on Data Ethics, Governance, and Quality in a Changing Data Ecosystem

Tuesday, 2nd September 2025

Outline

- I. Background
- II. Why Innovative Data?
- III. Traditional vs Innovative Data
- IV. Opportunities Presented by Innovative Data
- V. Challenges Associated with Innovative Data Sources
- VI. Strategies for Overcoming Challenges

Background

- Kenya, through the Kenya National Bureau of Statistics (KNBS), has made significant strides in diversifying the Country's data ecosystem.
- The emerging environment is characterized by administrative data systems, Earth Observation (EO), digital platforms, big data, AI and citizen-generated data.
- While these sources present opportunities, they also raise complex challenges relating to ethics, governance, quality, access, and sustainability.

Conti...

- While they can greatly improve decision-making, planning, and innovation, they require thoughtful integration, skilled human capital, and robust governance frameworks.
- National Statistical Offices and other organizations must modernize responsibly to harness these benefits while maintaining public trust and data integrity.

Conti....

Policy and Legal Anchoring:

- **Statistics Act, CAP 112 (Rev.2022):** KNBS is mandated to coordinate the National Statistical System and guarantee the quality, standards, consistency, and integrity of statistics.
- The Act also guarantees confidentiality in the production and dissemination of statistics
- **Data Protection Act, 2019:** Provides the legal framework for privacy, security, and ethical handling of personal data.

Why Innovative Data Sources?

- Increased demand for near real time, detailed, and disaggregated data. This necessitates the need to build a broad data ecosystem.
- Declining response rates and respondents fatigue in traditional surveys.
- Limited resources for the production of statistics using traditional sources.

Traditional vs Innovative Data

Traditional Data Sources:

- High cost
- Low frequency
- Limited coverage and granularity

Innovative Data Sources:

- Real-time availability
- High granularity
- Greater geographic coverage

Opportunities Presented by Innovative Data

- **Available expertise:** Because of emerging expertise like data science, machine learning, data architects
- **Technological advancements:** That enable faster data analysis, automated processes and enhanced data visualization
- **Enhanced Scope and Frequency:** Improves the timeliness of official statistics.
- **Cost Efficiency:** Reduces reliance on costly traditional surveys.
- **Public-Private Partnerships:** Supports collaboration for improved methodologies and access.

Challenges Associated with Innovative Data Sources

- **Data Quality and Incompleteness:** Issues with accuracy, noise, and missing data.
- **Complexity and Heterogeneity:** Difficulties in integrating unstructured and varied data formats.
- **Low standard:** Analytical tools that cannot handle high data volume and complexity.
- **Data Integration:** Linking multiple data sources requires technical capacity.
- **Timeliness and Retrieval:** Managing and organizing real-time data streams poses a logistical challenge.

Conti....

- **Skills and Talent Gaps:** Relevant expertise, such as programming language and tools for data linkage.
- **Infrastructure Limitations:** High demands on storage, processing, and cloud-based systems.
- **Data Privacy and Security:** Confidentiality and ethical issues.
- **Governance and Standards:** Lack of standardized rules for handling and interpreting data.
- **Policy and Regulation Gaps:** Lack clear legal frameworks to guide digital transformation.
- **Reputational Risks:** Misuse of data can erode public trust.

Strategies for Overcoming Challenges

A. Strengthening Data Governance

- Establish clear data ownership, responsibilities, and standards
- Implement policies for ethical data use and privacy compliance

B. Enhancing Data Quality and Integration

- Regularly validate and cleanse data to eliminate errors
- Integrate data from traditional and innovative sources to enrich insights
- Adopt automated tools for data management and transformation

C. Investing in Infrastructure and Skills

- Upgrade infrastructure to support real-time processing and big data storage
- Build internal data science capabilities through training and hiring

D. Fostering Collaboration

- Engage in partnerships (public-public, public-private) to gain access to data
- Work with Non State Actors for knowledge exchange

Thank you



www.knbs.or.ke

