REPORT ON ASSESSMENT OF GAPS AND OPPORTUNITIES FOR USING BIG DATA IN TANZANIA

September 2025

Republic of Tanzania The United

National Bureau of Statistics

Ministry of Finance

Dodoma

Office of the Chief Government Statistician

Presidents' Office – Finance and Planning

Zanzibar

The Report on Assessment of Gaps and Opportunities for using Big Data in Tanzania was produced by the National Bureau of Statistics (NBS) and Office of Chief Government Statistician (OCGS) in consultation with key stakeholders from a wide range spectrum including staff from Ministries, Departments, Agencies, Non-State actors, Research and Academic Institutions. This is the first comprehensive assessment of its kind to be conducted in Tanzania. The development of the assessment report was funded by the Government of Tanzania through the Tanzania Statistical Master Plan II (TMSP II) funds.

Additional information about the Report on Assessment of Gaps and Opportunities for using Big Data may be obtained from the National Bureau of Statistics, Head Office - Takwimu House, 64 Lusinde Road, P. O. Box 2683, 41104 Tambukareli – Dodoma, Tanzania. Tel: +255 26 – 2963822 E-mail: sg@nbs.go.tz; Website: www.nbs.go.tz: Office of Chief Government Statistician, P. O. Box 2321, 7 Fumba Road, 71125 Mjini Magharibi – Zanzibar. Tel: +255 24 – 2240134, E-mail: info@ocgs.go.tz; Website: www.ocgs.go.tz.

Recommended citation:

National Bureau of Statistics (NBS), and Office of Chief Government Statistician (OCGS). 2025. *The Report on Assessment of Gaps and Opportunities for using Big Data in Tanzania*, Dodoma, Tanzania: NBS and OCGS.

Contents

List of Figures

Abbreviations And Acronyms

CGD	Citizen-Generated Data
COSTECH	Tanzania Commission for Science and Technology
DHS	Demographic and Health Survey
FAO	Food and Agriculture Organization
FDES	Framework for the Development of Environment Statistics
GEE	Google Earth Engine
GEOSS	Group of Earth Observations System of Systems
HDX	Humanitarian Data Exchange
IoT	Internet of Things
NBS	National Bureau of Statistics
OCGS	Office of Chief Government Statistician
SDGs	Sustainable Development Goals
SEEA	System of Environmental-Economic Accounting
TMSP II	Tanzania Statistical Master Plan II
UN	United Nations

Preface

Tanzania stands at a pivotal moment in its development journey. As we strive to meet the ambitions of our national development plans and the global Sustainable Development Goals (SDGs), the demand for timely, granular, and reliable data has been at the fore front. Traditional data sources in terms of censuses and surveys have been the bedrock of our statistical system, but they are no longer sufficient alone. They are costly, infrequent, and often run short of the real-time insights needed for agile policymaking in a rapidly changing world.

This report, the first of its kind in Tanzania, is not merely an assessment; it is a call to action. Led by the National Bureau of Statistics (NBS) and the Office of Chief Government Statistician (OCGS) in close collaboration with key stakeholders from Ministries, Non-State Actors, Academia and research institutions in Tanzania. It represents a collective recognition that our nation's progress depends on the ability to harness the power of new data frontiers. Big Data from satellites, mobile phones, and digital transactions offers an unprecedented opportunity to complement the traditional methods, providing real time insights into poverty, agricultural productivity, environmental change, and more.

The findings within this document clearly reveals that, while we have the necessary infrastructure and the will to begin this transformation, we are held back by significant gaps in skills, policy, and coordination. The report provides a clear-eyed diagnosis of these challenges and a vision for a data driven future. It is intended to catalyze a national conversation and mobilize stakeholders across government, private sector, academia, and development partners to build a modern, integrated, and innovative National Statistical System for Tanzania.

Dr. Amina S. Msengwa
Statistician General,
National Bureau of Statistics
Dodoma.

Salum Kassim Ali
Chief Government Statistician,
Office of the Chief Government Statistician
Zanzibar

CHAPTER 1
INTRODUCTION

1.1 Background

The 2030 Agenda for Sustainable Development, with its core pledge to leave no one behind, has fundamentally altered the data needs of nations. Policymakers no longer require only national averages; they demand disaggregated, frequent, and actionable data to target interventions effectively. Globally, a revolution is underway and indeed, data has become an important resource, fueling innovation, governance, economic growth, and inclusive development, with Big Data and advanced analytics driving decision-making in real-time. Consequently, there is now a strong momentum towards the production of poverty, agriculture, environment and other statistics in different spectrums. This has resulted in an unprecedented demand for data under the various aspects, including data that are disaggregated according to the various characteristics relevant to assessing progress towards national, regional and international goals.

Tanzania, in line with global trends, acknowledges the urgency to harness data effectively to achieve sustainable socio-economic growth, improve governance, and enhance public accountability.

Currently, the government of Tanzania, through the National Bureau of Statistics (NBS) and Office of the Chief Government Statistician (OCGS), is implementing the second phase of the Tanzania Statistical Master Plan (TSMP II - 2022/23 - 2026/27). The TSMP II has given special attention to administrative data and data emerging from non-traditional sources such as geospatial or earth observation and citizen generated data as compliment to other data produced from censuses and surveys. This has partly been due to rapid use of data beyond traditional administrative and survey-based sources. The rise of big data which is generated from digital platforms, transactions, Internet of Things (IoT), and sensors, combined with Citizen-Generated Data (CGD) from communities, social media, mobile phone, other telecom location data and participatory initiatives, offers new opportunities for decision-making and policy formulation. Indeed, the Tanzania Statistical Master Plan II (TSMP II) explicitly recognizes this shift, highlighting the need to integrate non-traditional data sources like Big Data and

Citizen-Generated Data. Harnessing these diverse sources requires a well-established report on Assessment of gaps and opportunities for using big data and new technologies in Tanzania.

The persistence for this assessment report in Tanzania is amplified by rapid digital transformation, growing public and governance demands, and the availability of these new data sources such as administrative records, geospatial information and citizen generated data. These non-traditional data sources present significant opportunities to enhance data production. Importantly, this report is not treated in isolation; it is the direct product and logical evolution of the TSMP II. While TSMP II provides the operational roadmap for strengthening the National Statistical System (NSS), the assessment report serves as the overarching, integrative tool that empowers and expands upon TSMP II objectives. It moves beyond the production of statistics using traditional data sources (censuses and surveys) to accommodate and complement the non-traditional data sources or actors from the private sector and civil society, and break down silos and foster a unified, ethical, and innovative data production for Tanzania.

The assessment report is designed to act as a catalyst for innovation, transparency, and inclusive development, aligning closely with Tanzania's broader national development goals. This document will address existing gaps, leverage new data technologies, and strengthen data production to ensure all stakeholders, including the government, private sector, and civil society, can access reliable, timely, and actionable data.

The necessity of assessing gaps and opportunities for using big data for wide range of statistics including environment, poverty, health, trade, financial services, telecommunication and agriculture among others arises from an increasing demand for production of real-time and accurate data to inform policy-making, enhance service delivery, and enable proactive governance. Technological advancements are transforming data generation, analytics, and usage, making it imperative for NBS, OCGS and other data producers in Tanzania to adopt a structured and coordinated approach to manage this valuable resource.

Linking global initiatives on poverty, agriculture, health, trade, tax and environment statistics from a pool of statistics required under the Sustainable Development Goals (SDGs) and other international frameworks, to the Tanzania's national efforts involves a strategic alignment of overarching objectives with specific local actions.

This report argues that for Tanzania to achieve its socio-economic goals, it must bridge the gap between its robust traditional statistical system and the dynamic potential of new data sources. The persistence of data silos, infrequent data collection cycles, and limited granularity is no longer tenable. This assessment is the crucial first step in building a unified, responsive, and cost-effective data ecosystem that can keep pace with the demands of modern governance.

1.2 Purpose of the Assessment Report

The purpose of this assessment is threefold:

i. To critically identify the gaps in skills, infrastructure, policy, and awareness that currently limit the use of Big Data in Tanzania.

ii. To map the significant opportunities that these new technologies present, particularly for enhancing statistics in key sectors like poverty, agriculture, environment and more.

iii. To align with the implementation of the Disbursement Linked Indicators (DLIs) under the Tanzania Statistical Master Plan II (TSMP II).

1.3 Methods

The methodology combined a comprehensive desk review of key sectors with structured stakeholder interviews, engaging experts from government, academia, the private sector, and NGOs. This mixed-method approach ensures that the findings are grounded in both documentary evidence and the practical realities faced by data producers and users in Tanzania.

The details of two sources of information are as follow;

a. Desk review involves the review of relevant country documents from the selected three areas on poverty, agriculture and environment statistics in Tanzania as showcases; other documents; donor project documents; reports in the scientific literature; and web-based reports.
b. Interviews with key stakeholder groups and team of experts that included technocrats from NBS, OCGS, Ministry of Finance, Ministries responsible for Agriculture and environment, government agencies, academia, research institutions, financial services and telecommunication institutions, private sector representatives, members of Tanzania's Statisticians (TASTA), Non-State actors and staff of organizations working

in each of the three areas of focus. Stakeholders' questionnaire with structured modules and questions was prepared to achieve this objective.

1.4 Structure of the Assessment Report

The report is structured as follows;

i. Chapter 1 provides a snapshot of importance of the report and the role of data for policymaking; it states the objectives of the document and the methodology used to develop the assessment report;

ii. Chapter 2 explores the situation analysis on selected three areas as showcases using the desk review and stakeholders' interviews; and presents the benefits, best practices and Use Cases scenario from the selected countries regarding Integration of Big Data and new Technologies in different National Programs; and

iii. Chapter 3 provides a snap shot of challenges and Opportunities of using Big Data and new technologies in Tanzania.

References are provided to refer readers for further and in-depth reading around specific topics.

1.5 Target Audience

This document is intended to serve as a living document for NBS, OCGS, other data producers in Tanzania and all stakeholders interested in advocating for or investing in the management and use of Big Data and new technologies in enhancing the production of statistics with the objective to ensure a long term, sustainable, integration of geospatial data and Big Data technologies in Tanzania as far as statistics is concerned.

CHAPTER 2

CURRENT STATE AND APPLICATION OF BIG DATA IN TANZANIA

2.1 Overview

This chapter presents the background to the production of official statistics in Tanzania by highlighting traditional data sources and the historical conduct of household surveys and censuses. The analysis focuses on three selected areas: poverty, environment, and agriculture statistics as showcases, though the principles can be extended to other types of statistics. The chapter examines responses from stakeholders' interviews conducted using structured questionnaires and concludes with selected country case studies demonstrating successful applications of Big Data and new technologies in national programs that could be adapted to the Tanzanian context.

2.2 The Imperative for Big Data Integration

2.2.1 The Global Data Demand

The Sustainable Development Goals (SDGs) of the United Nations, particularly SDGs 2, 3, and 17, emphasize the critical importance of information sharing. However, despite this global call to action, many stakeholders in Tanzania and elsewhere lack actionable data-driven insights and a clear understanding of how data translates into meaningful action. High-quality data is paramount, as the costs of poor-quality data may exceed those of having no data at all (Cai and Zhu, 2015). Significant knowledge gaps persist regarding the complex linkages among agriculture, environmental issues, and poverty throughout the value chain. The transformative power of data to bridge these gaps remains largely untapped in Tanzania.

This global context directly influences Tanzania's data landscape. The country's commitment to the 2030 Agenda necessitates robust data systems capable of tracking progress, identifying gaps, and informing evidence-based policies. Traditional data collection methods, while valuable, are increasingly insufficient to meet these demands, creating an urgent need for innovative approaches. In what follows, three show case sectors are highlighted.

2.2.2 Environment Statistics: Framework and Applications

Guiding Frameworks

The production of environment statistics in Tanzania is guided by two complementary international frameworks. The Framework for the Development of Environment Statistics (FDES 2013) provides an organized structure for the scope, collection, and compilation of environment statistics. It outlines environmental statistics and offers a conceptual framework for synthesizing and categorizing data from various sources to support analysis, policy, and decision-making. The framework employs a tiered mechanism allowing countries to prioritize environmental data collection and methodological development according to their individual circumstances at national, regional, and local levels.

Complementing FDES, the System of Environmental-Economic Accounting (SEEA) provides guidance on integrating economic and environmental data into a single accounting framework consistent with the System of National Accounts. SEEA covers critical topics including agriculture, forestry and fisheries, air emissions, energy, water, land, environmental activity, and material flows. Together, these frameworks support reporting on several environment-related Sustainable Development Goals.

i. The Policy Imperative

a. The demand for environment statistics is increasing in tandem with the continued environmental challenges faced by modern society. The recognition that human well-being fundamentally depends on environmental health has led to increased emphasis on environmental and sustainability concerns requiring informed decisions and actions. Central to these actions is the regular production of high-quality environment statistics to support evidence-based policymaking by enabling the identification of environmental policy issues and allowing their objective quantification.

b. Environment statistics portray key information about the state of the environment and its most relevant changes through space and time. These statistics are necessary for producing environmental assessments, state of the environment reports, environmental compendia, environmental indicators, indicators of sustainable development, and for facilitating environmental economic accounting.

ii. Current Data Sources and Limitations

In Tanzania, environmental statistics are often compiled using administrative data sources, such as natural resources, land records and climate data from the Tanzania Meteorological Authority (TMA). The production of environment statistics also relies on traditional data sources such as censuses and surveys. While administrative and traditional data sources provide valuable insights, they frequently suffer from quality issues, time lags, high costs, and limited coverage.

Recognizing these limitations, a few institutions in Tanzania have begun utilizing non-traditional data sources to complement traditional environment statistics. However, this adoption remains limited and uncoordinated, highlighting the need for a more systematic approach to integrating Big Data and new technologies into environmental monitoring and reporting.

2.2.3 Agricultural Statistics: Foundation of Tanzania's Economy

i. Economic Significance

Agriculture forms the foundation of the Tanzanian economy, accounting for approximately half of national income and three-quarters of merchandise exports. It serves as a source of food and provides employment to about 80% of Tanzanians. Tanzanian agriculture is dominated by smallholder farmers cultivating farms of less than three hectares, relying primarily on rainfed rather than irrigated agriculture. About 70% of Tanzania's crop area is cultivated by hand hoe, 20% by ox plough, and 10% by tractor. Food crop production dominates the agriculture economy, though irrigated agriculture in some areas helps stabilize agricultural production, improve food security, increase farm productivity and income, and produce higher-value crops such as vegetables and flowers.

ii. Data Requirements and Challenges

Agricultural statistics provide essential information on crop production and yield, extent of arable land and sown area, livestock numbers, and prices of land and agricultural products. Traditional sources of agricultural and rural statistics include administrative reporting systems, household surveys, and censuses of agriculture.

However, the increasing demand for more granular, timely, accurate, and reliable data is pushing national statistical agencies to explore innovative and more efficient methods of collecting agricultural data.

The agricultural sector faces a novel set of challenges, including climate change, fluctuations in supply and demand, workforce disruptions, and supply chain challenges (FAO, IFAD and WFP, 2015). Big Data analytics in the agricultural sector has tremendous potential to address various food production requirements. Advanced practical and systematic strategies are needed to correlate different factors driving agriculture to derive valuable information. Leveraging Big Data in the agricultural sector can provide insights into farming practices, enable real-time decision-making, and motivate incorporation of new farming methods to enhance resilience and reduce potential risks (Wolfert et al., 2014).

While a few institutions in Tanzania have started utilizing non-traditional data sources to complement traditional agricultural statistics, systematic integration remains limited.

2.2.4 Poverty Statistics: Measuring Development Progress

i. Critical Role in Policy Making

Access to high-quality and timely information about poverty is essential for designing and targeting effective policies, strategies, and programs. Poverty statistics play a critical role in monitoring progress towards national and regional priorities, as well as global commitments such as the Sustainable Development Goals (SDGs).

Tanzania has made notable progress towards poverty reduction over the past two decades, though progress has slowed in recent years. Poverty declined marginally, reaching a headcount rate of 26.4 percent according to the 2017-18 Tanzania Mainland Household Budget Survey (HBS). Poverty is concentrated in rural areas, where 81 percent of the country's poor reside. In Zanzibar, the proportion of people living below the food poverty line declined from 10.8 percent to 9.3 percent between 2014/15 HBS and 2019/20 HBS, while basic needs poverty declined from 30.4 percent to 25.7 percent over the same period.

ii. Traditional Methods and Their Limitations

In Tanzania, poverty measurement has traditionally relied on Household Budget Surveys. These multi-purpose surveys are flexible in content and collect detailed information on key variables such as consumption, assets, and other multidimensional poverty indicators. While they provide valuable insights, these surveys face significant limitations: they are costly, infrequent (conducted every five years), and often published with substantial time lags. Furthermore, survey sample sizes allow reliable estimates only at national, location (rural/urban), and first-level subnational levels (regions or zones), but not at lower levels such as districts where policymakers also need insights.

iii. The Big Data Opportunity

While traditional methods remain essential, their limitations, high costs, long timelines associated with data collection and publication delays, and coarse granularity are driving interest in leveraging Big Data for poverty measurement to complement and enhance traditional approaches. With increased digitalization and exponential growth of data locally and globally, promising signs indicate that novel Big Data sources can help provide cost-effective, time-efficient, accurate, and up-to-date indicators (Steele et al., 2017).

2.3 Analysis of Stakeholders' Interviews

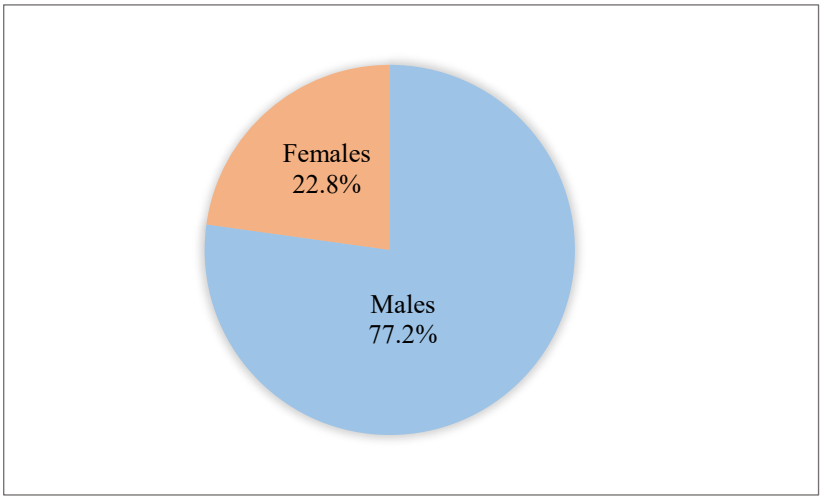
To understand the current landscape and challenges of Big Data adoption in Tanzania, the National Bureau of Statistics (NBS) and Office of Chief Government Statistician (OCGS) conducted a comprehensive stakeholders' survey entitled "Assessment of challenges and opportunities of using Big Data in Tanzania." The survey aimed to identify key challenges and opportunities regarding the application of Big Data and new technologies to complement traditional data sources. The structured questionnaire, developed using Survey Solutions (SuSo) Software, was distributed to respondents from a wide range of stakeholders.

2.3.1 Characteristics of Respondents

A total of 202 respondents participated in the interviews. The gender distribution of respondents showed 156 males (77.2%) and 46 females (22.8%), Figure 2.1. Respondents represented diverse sectors, with the majority from policy and planning, as well as academic and research institutions (23.2% each), followed by education (14.7%), agriculture (14.1%), health (7.9%),

environment (5.9%), financial services (4.7%), poverty (2.6%), telecommunication and mobile services (1.9%), and transport (1.8%) sectors. This diverse representation ensures that the findings reflect perspectives across Tanzania's key economic and social sectors.

Figure 2.1: Percent Distribution of Surveyed Respondents by Sex



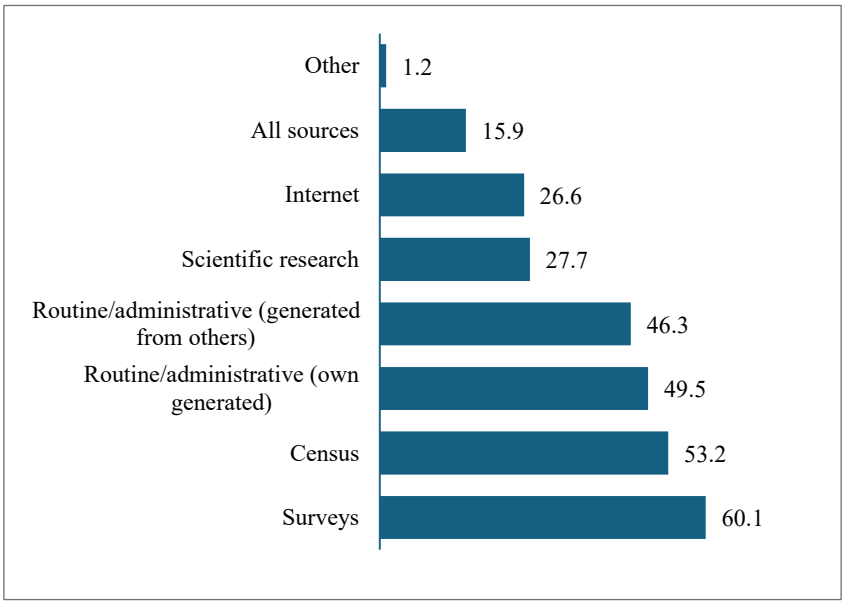
2.3.2 Challenges

i. Current Use of Big Data and New Technologies

The survey revealed that the use of Big Data and new technologies within sectors remains uncommon in Tanzania. When asked about data sources commonly used in their workplaces, respondents indicated a heavy reliance on traditional methods. Surveys were the most common data source used by 60.1% of respondents, followed by census data (53.2%) and own-produced routine or administrative data (49.4%), Figure 2.2. This pattern confirms that despite growing global trends toward Big Data adoption, Tanzania's institutions continue to depend primarily on conventional data collection approaches.

This finding is particularly significant as it establishes a baseline for understanding the digital transformation challenge facing Tanzania. The low adoption of Big Data and new technologies indicates substantial room for growth but also suggests that significant capacity building and infrastructure development will be required.

Figure 2.2: Percentage of Types of Data Sources Commonly Used

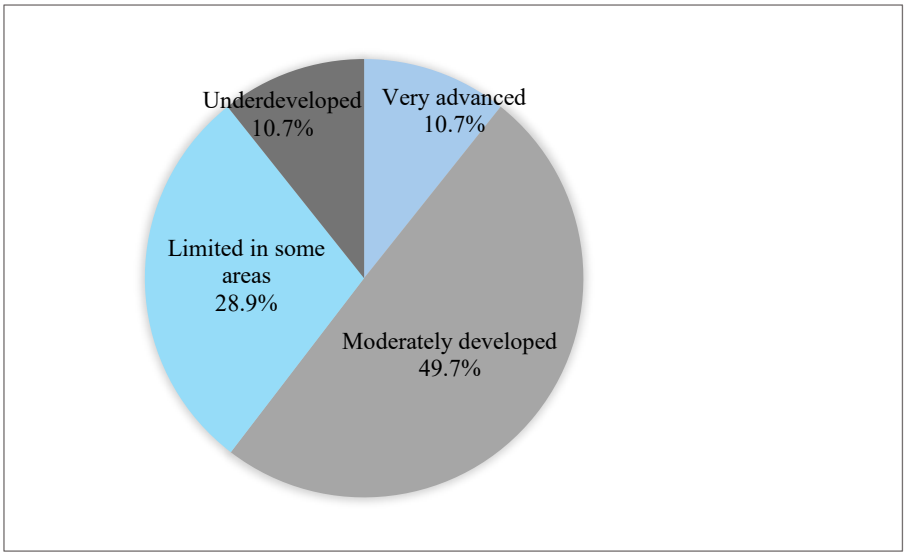


ii. Current State of Statistical Infrastructure

Respondents were asked to rate the current state of statistical and digital infrastructure available in their institutions. The results revealed moderate development at best. Only 10.7 percent of respondents reported that their institutions have very advanced statistical digital infrastructures. Approximately half (49.7%) indicated that their institutions have moderately developed statistical digital infrastructure, while the remaining respondents reported either basic or underdeveloped infrastructure, Figure 2.3.

This infrastructure gap represents a fundamental challenge to Big Data adoption. Without adequate technological foundations, even institutions with interest and potential use cases for Big Data cannot effectively implement solutions. The moderate infrastructure development suggests that while some progress has been made, substantial investment is needed to create an enabling environment for widespread Big Data utilization.

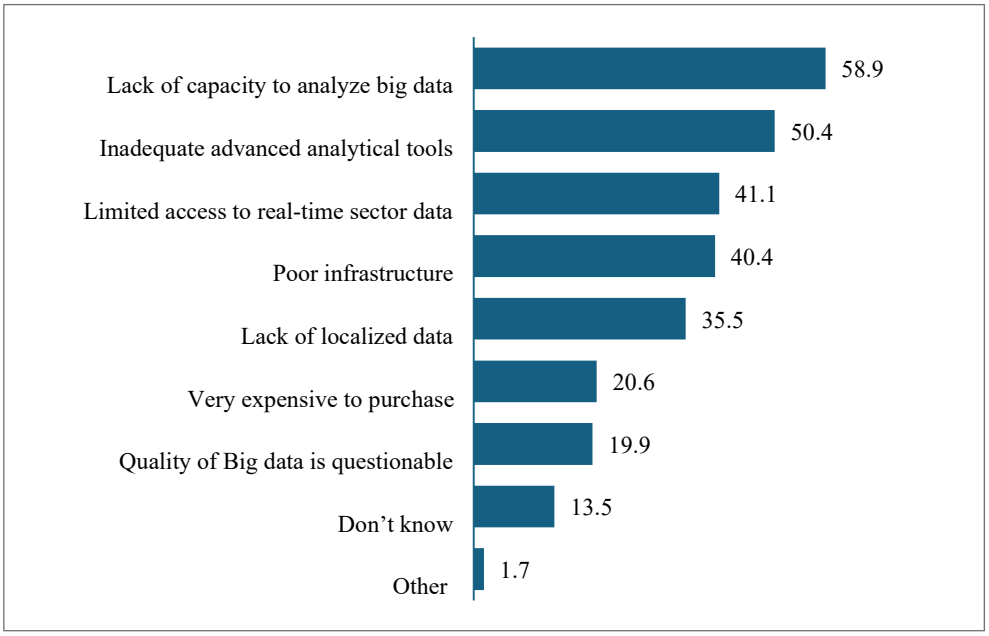
Figure 2.3: Percent Distribution of Level of Statistical Infrastructure



iii. Usage of Big Data and New Technologies

The survey identified several interconnected challenges hindering Big Data adoption in Tanzania. Among respondents, 58.9 percent identified lack of capacity to analyze Big Data as a high challenge, followed by inadequate advanced analytical tools (50.4%), Figure 2.4. Only a small proportion (13.5%) reported not knowing about the challenges of using Big Data, suggesting relatively good awareness of the issues even if solutions remain elusive.

Figure 2.4: Percent of Challenges in Using Big Data and New Technologies

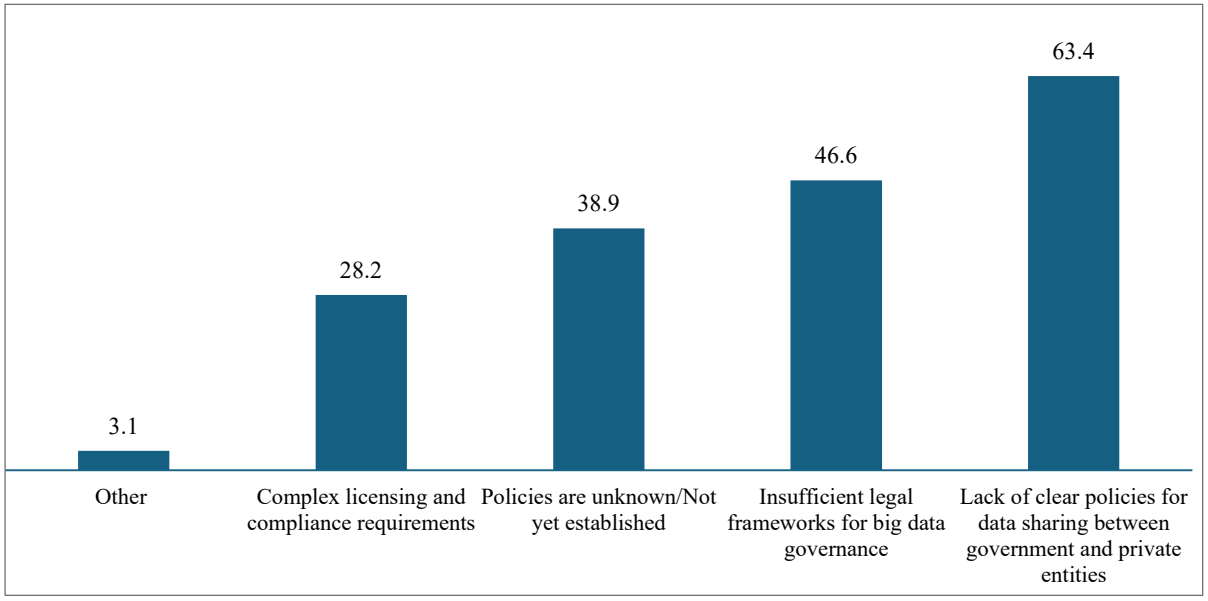


iv. Policy and Regulatory Challenges

Respondents were further asked about specific challenges related to policy and regulation in accessing and using Big Data at their institutions. The majority (63.4%) reported lack of clear policies for data sharing between key entities as a major concern. This challenge is particularly significant because data sharing is fundamental to realizing the full potential of Big Data analytics. Without clear frameworks for how data can be shared, accessed, and used across institutions, even available datasets remain siloed and underutilized.

Additionally, 46.6 percent of respondents reported limited legal frameworks for Big Data governance, and 38.9 percent indicated that policies are either unknown or have not been established, Figure 2.5. This policy vacuum creates uncertainty and risk aversion among institutions that might otherwise explore Big Data applications. The absence of clear governance frameworks also raises concerns about data privacy, security, and ethical use.

Figure 2.5: Percent of Challenges Related to Policies and Regulations in Using Big Data and New Technologies



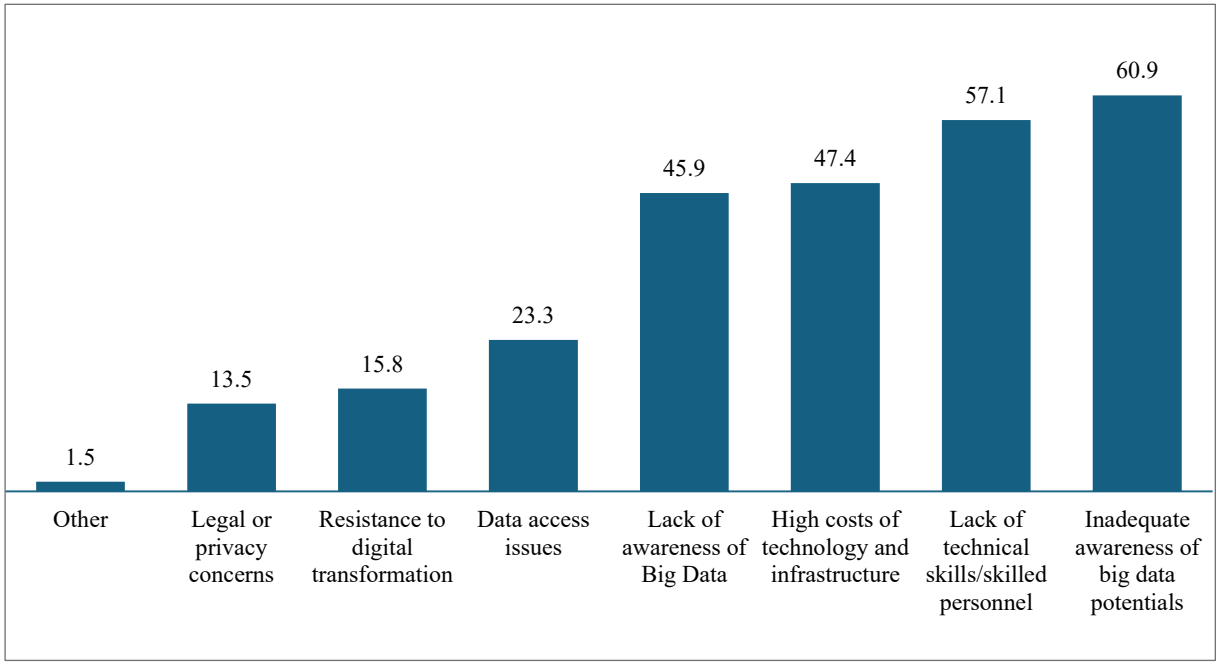
v. Adoption-Related Challenges

Respondents also reported significant challenges regarding the adoption of Big Data and new technologies. The majority (60.9%) cited inadequate awareness of Big Data potentials as a barrier. This finding suggests that many stakeholders do not fully understand what Big Data can accomplish or how it might benefit their specific contexts. Following closely, 57.1%

identified lack of technical skills and skilled personnel as a challenge, while 47.4% pointed to high technological and infrastructure costs, Figure 2.6.

Interestingly, only 15.8 percent reported resistance to digital transformation, suggesting that the primary barriers are practical rather than cultural or psychological. Stakeholders appear willing to adopt new technologies if appropriate support, skills, and resources are available.

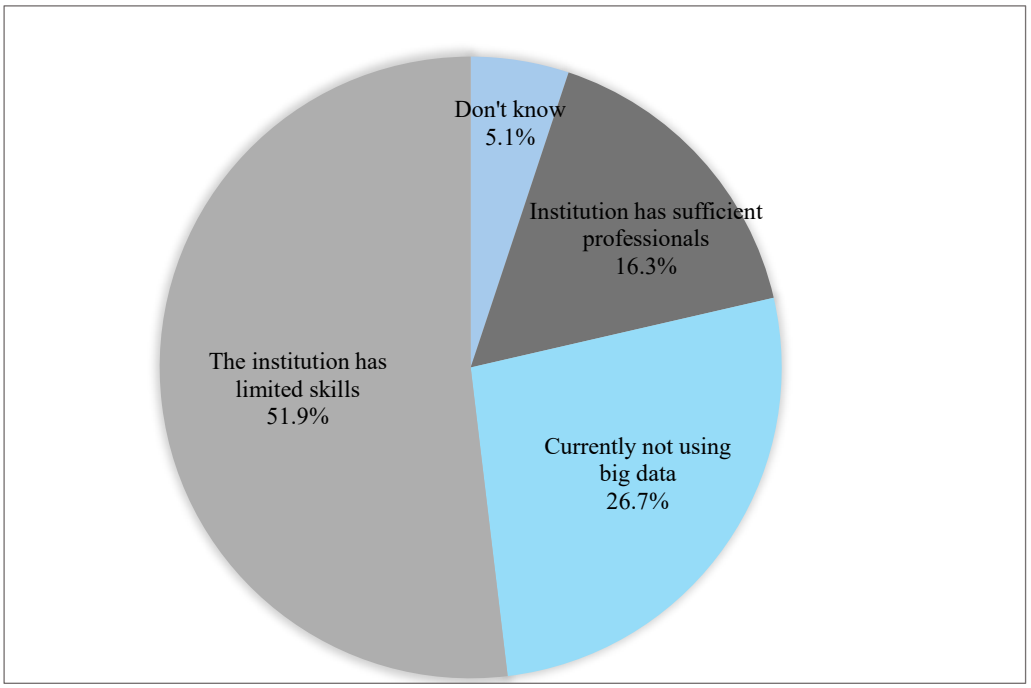
Figure 2.6: Percent of Challenges Related to Adoption in Using Big Data and New Technologies



vi. Skills and Capacity Assessment

The skills gap emerged as one of the most critical challenges facing Big Data adoption in Tanzania. When asked to rate their institutions' skills and capacity in utilizing Big Data and new technologies, only 16.3 percent of respondents indicated that their institutions have sufficient professionals. More than half (51.9%) reported that their institutions have staff with limited skills, while more than a quarter (26.7%) stated they are currently not using Big Data at all (Fig 2.7). This skills deficit creates a self-reinforcing cycle: without skilled personnel, institutions cannot effectively implement Big Data solutions, and without practical experience using Big Data, staff cannot develop relevant expertise. Breaking this cycle requires targeted interventions in capacity building and human resource development.

Figure 2.7: Percent Distribution of Level of Skill and Capacity

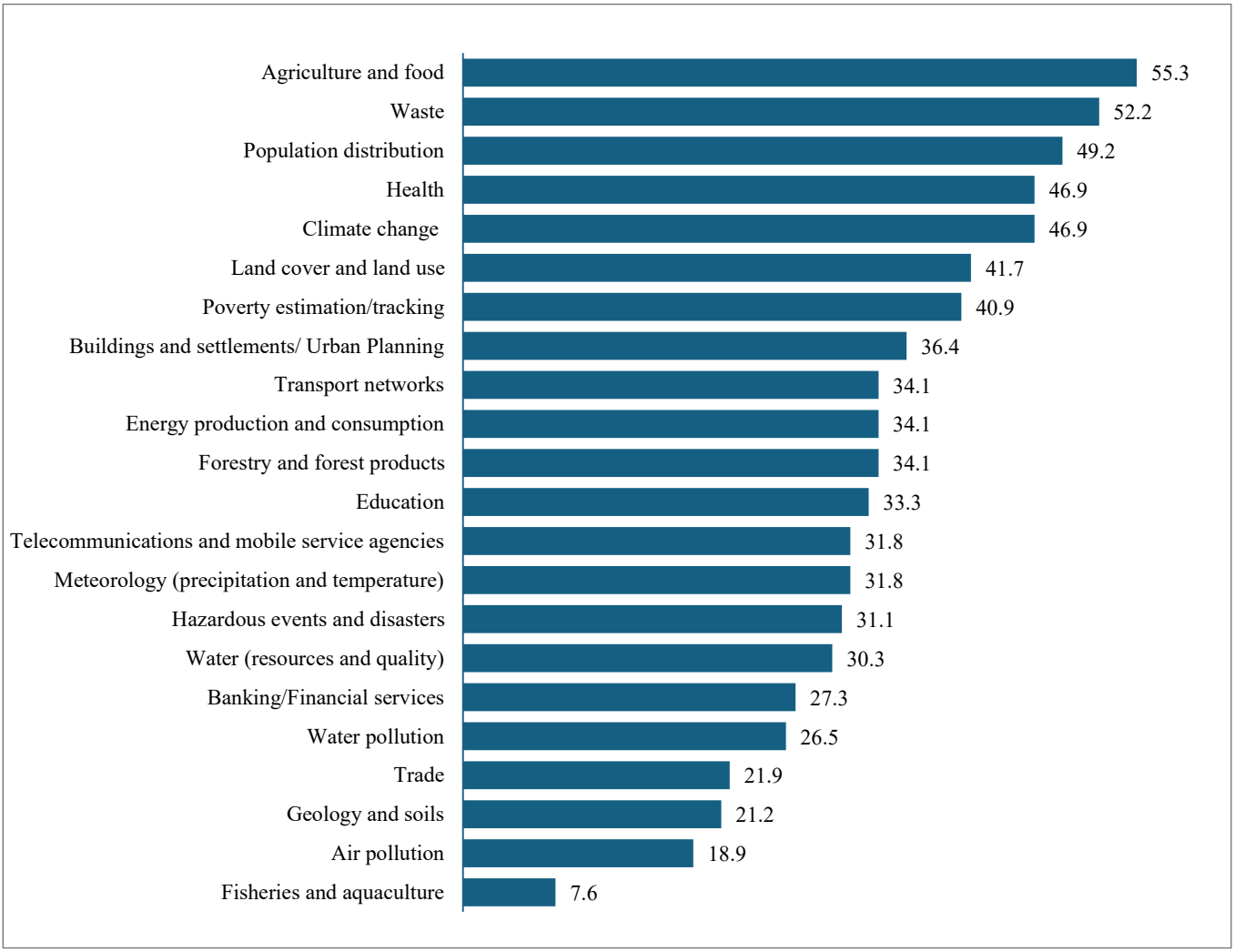


2.3.3 Opportunities and Priority Areas

Despite the challenges, respondents identified significant opportunities for Big Data application. When asked which sectors could benefit from Big Data and new technologies in Tanzania, more than half identified agriculture and waste management as priority topics. Other sectors mentioned included health, education, environment, financial services, poverty monitoring, and transport, Figure 2.8. Notably, less than ten percent of respondents mentioned fisheries and aquaculture (7.6%), despite Tanzania's significant coastline and inland water resources, suggesting potential in ability in opportunity identification.

These findings indicate that stakeholders recognize Big Data's potential value and can identify relevant applications within their sectors. The challenge lies not in recognizing opportunities but in building the capacity to realize them.

Figure 2.8: Percent of Proposed Key Areas of Interest that Could Benefit from Using Big Data and New Technologies



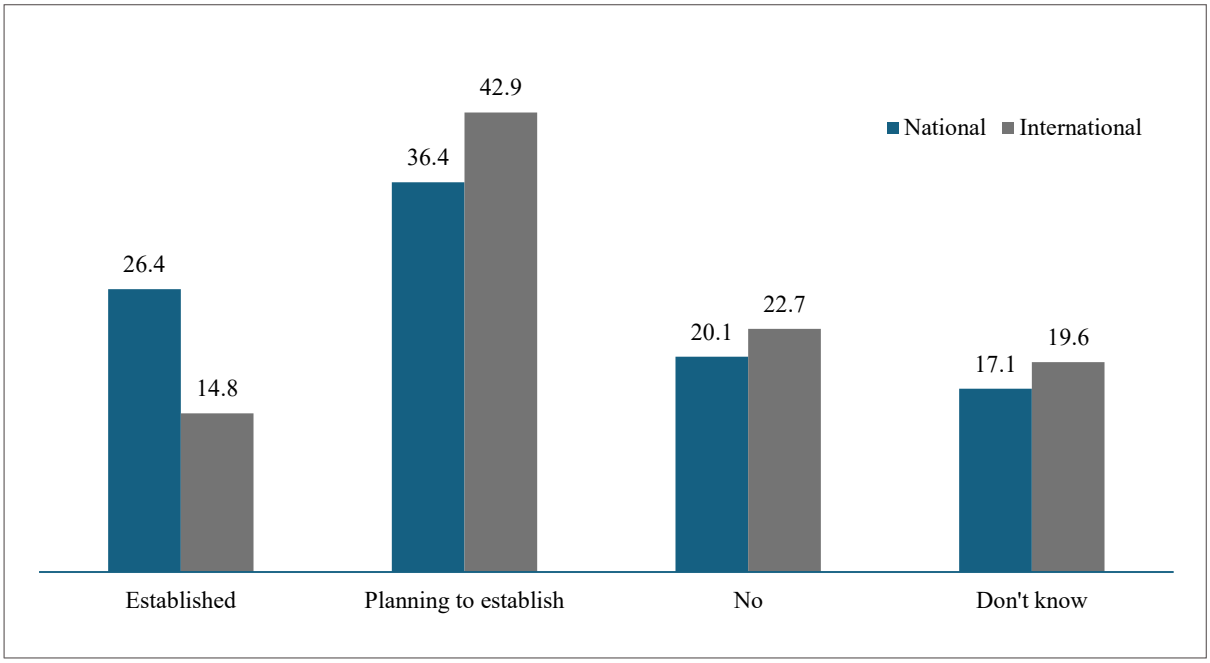
i. Partnership Initiatives at National and International Level

Collaboration emerged as a key theme in the survey responses. When asked about partnerships related to Big Data and new technologies, more than 4 in 10 respondents (42.9%) reported that their institutions are planning to establish partnerships with relevant organizations at the international level, while 36.4 percent indicated plans for national-level partnerships. These figures suggest growing recognition that Big Data challenges require collaborative solutions.

Regarding existing partnerships, 26.4 percent of respondents reported that their institutions have established partnerships with other organizations in Tanzania, compared to 14.8 percent (Figure 2.9) who reported partnerships with organizations outside Tanzania. The relatively low

rates of established partnerships, compared to planned partnerships, suggest that Tanzania is in the early stages of building the collaborative networks needed for effective Big Data utilization.

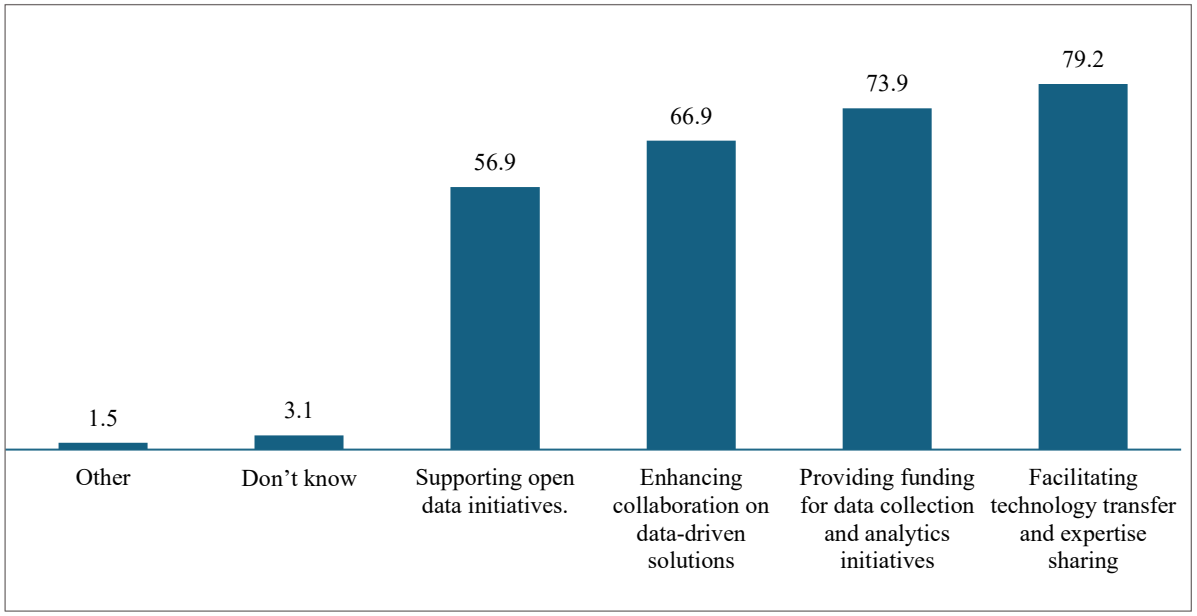
Figure 2.9: Percent of Respondents Who Reported Status of Partnership About Big Data Initiatives



ii. Roles for Public-Private Partnerships

Respondents provided valuable insights into potential roles for public-private partnerships in advancing Big Data adoption. More than 7 in 10 suggested technology transfer and knowledge sharing as critical functions, along with support for data collection and analytics. More than half (56.9%) emphasized supporting open data initiatives (Figure 2.10). These responses indicate that stakeholders view partnerships not merely as funding mechanisms but as vehicles for knowledge exchange, capacity building, and infrastructure development.

Figure 2.10: Percent of Respondents Who Suggested the Roles of Public-Private Partnerships in Advancing the Use of Big Data

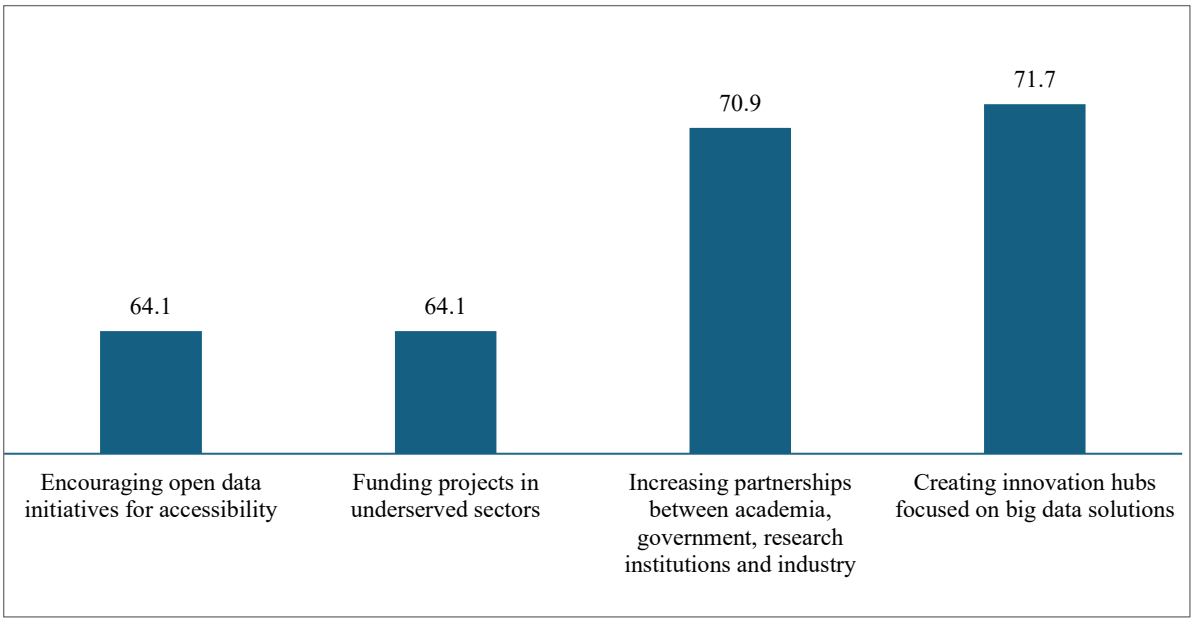


iii. Suggested Initiatives for Promotion

Respondents offered concrete suggestions for promoting Big Data and new technologies in Tanzania. More than 7 in 10 respondents suggested establishing innovation hubs focused on Big Data solutions and increasing partnerships between key players (Fig.2.11). These recommendations align with successful approaches used in other countries and reflect understanding that Big Data adoption requires ecosystem development rather than isolated interventions.

Other frequently mentioned initiatives included capacity building programs, policy development, infrastructure investment, and awareness campaigns. The consistency and practicality of these suggestions indicate that stakeholders have thoughtfully considered what would enable Big Data adoption in their contexts.

Figure 2.11: Percent of Proposed Initiatives to Promote use of Big Data and New Technologies



2.4 International Best Practices: Country Case Studies

Examining successful Big Data applications in other countries provides valuable insights for Tanzania's development pathway. The following case studies demonstrate benefits and best practices from various national programs that could be adapted to Tanzania's context.

i. Low-cost District Mapping using Community Health Workers in Rwanda

The University of Rwanda demonstrated an innovative approach to geospatial data acquisition by engaging Community Health Workers (CHWs) to collect data using GPS receivers. By developing GPS training materials in the local language and adopting participatory methods, the initiative achieved a 50% reduction in geospatial data acquisition costs compared to using GIS professionals.

The resulting information provided accurate mapping of physical accessibility to health services, improved scheduling of outreach vaccination services, and generated evidence-based advocacy that led to additional health resources. Success factors included participatory approaches involving local workforce, production of low-literacy training materials, and significant collaboration and data sharing between the University of Rwanda, Ministry of Health, and National Statistics Office.

Relevance to Tanzania: This case demonstrates that Big Data solutions need not be expensive or require highly specialized personnel. Tanzania could adopt similar community-based approaches, leveraging existing health worker networks and adapting training materials to local languages and literacy levels. The collaborative model involving universities, government ministries, and statistical offices mirrors Tanzania's institutional structure.

ii. Experimental sugarcane statistics with non-survey data in Australia

The Australian Bureau of Statistics (ABS) developed experimental sugarcane statistics using non-survey data and methods, potentially reducing reporting burden on farmers. Collaborating with agricultural experts, ABS integrated existing data from administrative sources like the Australian Sugar Milling Council Levy Payer Register and satellite imagery with traditional survey data to create more detailed and timely regional statistics than traditional surveys provide.

This approach enhanced data quality by providing insights into local areas, enabling better understanding of issues like natural disaster impacts, and supporting development of new methods applicable to other agricultural commodities.

Relevance to Tanzania: Given agriculture's central role in Tanzania's economy and the dominance of smallholder farming, reducing farmer reporting burden while improving data quality offers significant benefits. Tanzania could explore similar integration of administrative data from agricultural levy systems, cooperative records, and satellite imagery to complement traditional agricultural surveys.

iii. Tracking Vaccination Teams using GPS and Satellite Images in Nigeria

With support from the Bill and Melinda Gates Foundation, Nigeria applied innovative geospatial mapping technologies, including automated Feature Extraction from high-resolution satellite images and GPS tracking techniques, to support vaccination activities in 10 Northern States with incomplete, inaccurate, and outdated maps.

Interpretation of satellite images allowed precise location of remote settlements and hamlets, while GPS tracking improved efficiency in microplanning vaccination team assignments by enabling daily activity tracking, near-real-time supervision, and

corrective interventions to optimize coverage. The resulting detailed GIS maps, including all settlements, points of interest, secondary and tertiary roads, allowed accurate planning and monitoring of vaccination teams in previously unmapped areas. This improvement dramatically reduced chronically missed settlements from 4.1% to less than 0.05% in one year.

Relevance to Tanzania: Tanzania faces similar challenges in providing health services to remote and rural populations. This approach could be adapted to improve vaccination coverage, particularly in hard-to-reach areas, and could extend to other health services and social programs requiring field team management and monitoring.

iv. Identifying crop type at national level with Earth Observation Data in Senegal

Agriculture's strategic importance for socio-economic development and food security in Senegal mirrors Tanzania's situation. Policymakers required quality data available in near real-time to track agriculture, including crop type identification. However, traditional data collection methods proved too slow and expensive.

The Senegal Directorate of Analysis, Forecasting and Agricultural Statistics (DAPSA) partnered with the Food and Agriculture Organization (FAO) to apply Sentinel 2 satellite images, geo-referenced with Senegal's National Agricultural Census 2018, as inputs to machine learning models for producing national agriculture statistics. This approach leveraged unique spectral signatures of crops integrated with field data from sources like FAO's EOSTAT project to create accurate, granular crop type maps.

The joint work produced the Sen2Agri Software, and the DAPSA team acquired new Big Data skills and best practice methods for crop geo-referencing. DAPSA can now produce national maps of certain crops and extract statistics of national importance.

Relevance to Tanzania: This case demonstrates how national statistical offices can build internal capacity while developing practical Big Data applications. Tanzania could pursue similar partnerships with FAO or other international organizations, leveraging its own agricultural census data with satellite imagery to develop crop monitoring systems. The capacity building component is particularly relevant given Tanzania's identified skills gaps.

v. Improving Land Statistics with Earth Observation data in India.

India's Ministry of Statistics and Programme Implementation (MoSPI) is among few statistical institutions globally producing Environment Accounts using remote sensing data. To fulfil its mandate of issuing policy fit environmental accounts, MoSPI formed an Inter-Ministerial Group on Environmental Accounting comprising the Ministry of Environment, Forests and Climate Change, Ministry of Earth Sciences, National Remote Sensing Center, and Comptroller and Auditor General.

Earth Observation (EO) data improves India's land statistics by providing high-resolution, continuous monitoring of agriculture, land use, and environmental conditions, complementing traditional census data for detailed farm-level insights and national-level monitoring through systems like VEDAS.

Relevance to Tanzania: Tanzania's commitment to environmental accounting under SEEA could benefit from similar inter-ministerial coordination and systematic use of Earth Observation data. The inter-ministerial group model offers a governance structure that could address Tanzania's identified challenge of limited coordination across institutions.

vi. The use of Earth Observation (EO) Data for computing SDG Indicator 6.6.1

Some Sustainable Development Goals indicators can only be measured using Big Data sources due to lack of traditional data. For example, SDG Indicator 6.6.1 measures change in the extent of water-related ecosystems over time and distribution of inland open water bodies.

The UN Environment Programme partnered with the European Commission Joint Research Centre and Google to review and merge Big Data for mapping location and monitoring water surfaces, incorporating satellite imagery from the past 35 years. This collaboration produced the Global Surface Water Explorer (GSWE) tool, a free, online global dataset accessible to all stakeholders including research institutions, researchers, private companies, and National Statistical Offices. The UN Freshwater Ecosystems Explorer provides access to these global datasets and methodologies for countries to use for monitoring and decision-making.

Relevance to Tanzania: This case demonstrates how global Big Data initiatives can support national reporting obligations. Tanzania could leverage existing global platforms and datasets to report on SDG indicators that would otherwise require costly primary data collection. The open-access nature of these tools reduces barriers to entry for countries building Big Data capacity.

2.4.1 Synthesis of Lessons Learned

These case studies collectively demonstrate several key principles:

i. **Collaboration is Essential:** Successful Big Data initiatives involve partnerships between government agencies, international organizations, universities, and private sector.
ii. **Capacity Building Must Be Integrated:** Effective programs build local capacity rather than creating dependency on external expertise.
iii. **Solutions Can Be Cost-Effective:** Community-based approaches and open-source tools can significantly reduce costs compared to traditional methods.
iv. **Existing Data Can Be Leveraged:** Combining administrative data, satellite imagery, and traditional surveys yields better results than any single source.
v. **Global Resources Are Available:** International platforms and datasets can support national initiatives, particularly for SDG monitoring.
vi. **Start Small and Scale:** Successful programs often begin with pilot projects in specific sectors or regions before expanding nationally.

These lessons directly address many of the challenges identified in Tanzania's stakeholder survey and provide practical pathways forward. There is scope to apply these case studies to be adopted to Tanzania context in addition to the list of proposed indicators as presented in **Appendix A**.

CHAPTER 3

CHALLENGES AND OPPORTUNITIES FOR BIG DATA INTEGRATION IN TANZANIA

3.1 Challenges and Opportunities for Using Big Data and New Technologies in Tanzania

This chapter synthesizes the findings from Chapter 2 to present a comprehensive analysis of challenges and opportunities for using Big Data and new technologies in Tanzania. Drawing from international best practices, stakeholder perspectives, and current trends across sectors, this analysis provides the foundation for strategic recommendations to advance Tanzania's data-driven development agenda.

3.1.1 Critical Challenges Impeding Big Data Adoption

The main challenges in using Big Data and new technologies in Tanzania extend beyond technical issues to encompass institutional, human resource, and policy dimensions. These challenges, demonstrated through existing literature and stakeholder interview responses, are interconnected and mutually reinforcing, requiring comprehensive solutions rather than isolated interventions.

Institutional framework:

A sustainable and cost-effective use of Big Data and new technologies for decision-making across sectors requires a supportive institutional environment. Unfortunately, elements constituting such an environment including strategy and plans, governance structures, policies, and financial resources are often lacking or incomplete, preventing institutionalization and long-term sustainability.

The lack of familiarity and awareness regarding the use and benefits of Big Data as decision-making tools among researchers, policymakers, planners, program managers, and other stakeholders poses a major obstacle to introducing these technologies into organizational budgets and strategic plans. This awareness gap is compounded by vertical program structures and lack of inter-institutional collaboration and knowledge sharing, creating isolated pockets of activity rather than coordinated ecosystem development.

As stakeholder interviews revealed, 63.4 percent of respondents identified lack of clear policies for data sharing between key entities as a major challenge, while 46.6 percent reported limited legal frameworks for Big Data governance. This institutional vacuum creates uncertainty and inhibits investment in Big Data capabilities.

i. Limited skills and Human resources

Accessing, processing, and analyzing Big Data requires competent personnel with specialized or sophisticated skills. NBS, OCGS, and other data producers need staff capable of not only technical data analysis but also understanding the methodological implications of using Big Data for statistical production. Using Big Data may require new methods and techniques, including advanced modeling approaches, especially when producing early indicators or nowcasting. Artificial Intelligence and deep learning techniques may be necessary for processing unstructured text messages or satellite images.

The stakeholder survey confirmed this challenge, with 58.9 percent of respondents identifying lack of capacity to analyze Big Data as a high challenge, and 51.9 percent reporting that their institutions have staff with only limited skills. This skills deficit is not merely about technical training; it reflects deeper gaps in educational curricula, professional development opportunities, and practical experience with Big Data applications.

Furthermore, the challenge extends beyond individual skills to organizational capacity. Institutions need teams that combine data science, domain expertise, and policy understanding. Building such multidisciplinary teams requires sustained investment in recruitment, training, and retention strategies.

ii. Quality of Big Data

Big Data is often largely unstructured, meaning that such data sources have no pre-defined data model and do not fit well into conventional relational databases. The quality issue is mainly linked to the absence of agreed-upon and enforced guidelines, specifications, standards, and protocols across partners within various sectors.

These gaps result in data collected with different levels of accuracy that are not properly related or linked with data from other sources and are often poorly documented. Such inconsistencies make it difficult to ensure proper use and can potentially lead to errors impacting decision-

making. Additionally, the level of completeness and timeliness observed for some auxiliary Big Data relevant to sector programs remains an important issue.

Unlike traditional statistical data collected through carefully designed censuses and surveys with known quality parameters, Big Data often comes from sources designed for operational purposes rather than statistical analysis. Adapting such data for statistical production requires methodological rigor and quality assurance frameworks that are not yet well-established in Tanzania.

iii. Accessibility of Big Data

Challenges around data accessibility are primarily related to two issues: first, the difficulty of discovering which Big Data sets are available and where to find them; and second, restrictions on access to and use of available data. While growing efforts toward establishing partnerships among key players nationally and internationally aim to address these issues, coupled with development of online data catalogues and release of open data policies, enforcement remains difficult and often conflicts with other laws and regulations within institutions.

Private sector entities holding valuable data may be reluctant to share due to commercial sensitivities, lack of trust in data security measures, or absence of clear legal protections. Government agencies may face bureaucratic restrictions or inter-departmental rivalries that inhibit data sharing. Research institutions may lack mechanisms to access operational data from government or private sources.

The stakeholder survey revealed that only 26.4 percent of respondents reported established partnerships with other organizations in Tanzania for Big Data initiatives, suggesting that collaborative data ecosystems remain underdeveloped.

iv. Awareness Deficits

Low levels of awareness about potential uses and applications of Big Data and new technologies persist across various sectors in Tanzania. This challenge, evidenced by stakeholder interview responses with 60.9percent citing inadequate awareness of Big Data potentials, operates at multiple levels.

At the leadership level, decision-makers may be fully aware of how Big Data could improve their organization's effectiveness or inform better policies. At the practitioner level, staff may not recognize opportunities to apply Big Data approaches to their work. At the public level, citizens and stakeholders may not understand how data could empower them or improve service delivery.

This awareness gap is particularly problematic because it undermines demand for Big Data solutions. Without understanding potential benefits, organizations are unlikely to prioritize investments in data infrastructure, skills development, or partnership building.

v. Technical Capacity Gaps

While some institutions organize staff training on Big Data and new technologies, such training is generally more focused on using particular Big Data or geospatial solutions (e.g., GIS software or specific types of analysis) and rarely covers Big Data management practices or integration of geospatial data and technologies with other data sources from a broader perspective.

Moreover, these training events are typically one-off initiatives with little or no follow-up or updates on recent technological developments, resulting in difficulties maintaining consistent Big Data capacity in institutions. As technology evolves rapidly, skills can quickly become outdated without continuous learning and professional development.

The challenge extends beyond initial training to include lack of practical application opportunities. Even well-trained staff may lose skills without regular use, and institutions may lack the infrastructure or data to provide meaningful practice opportunities.

3.1.2 Strategic Opportunities and Enabling Trends

While substantial challenges exist for using Big Data and new technologies across different sectors, several initiatives in Tanzania and globally can help address them. These opportunities represent potential pathways for overcoming identified barriers and building a robust Big Data ecosystem supporting Tanzania's development objectives.

Global, Regional and National Institutional Frameworks

In recent years, significant efforts and resources have been devoted at global, regional, and national levels with common objectives of:

a. Recognizing the importance of strengthening the use of big data and new technologies,
b. Reaching a more collaborative and coordinated approach to big data management and use,
c. Improving the availability, quality and accessibility of big data developed and maintained by countries,
d. Establishing the necessary institutional framework to sustain the established capacities on the long term.

These initiatives include, from the global to the national level:

i. Global Frameworks

The Sustainable Development Goals (SDGs) and the Sendai Framework for Disaster Risk Reduction provide the landscape for developmental and humanitarian agendas and offer important leverage to support introduction and strengthening of Big Data and new technologies in health, environment, and related sectors in countries. These frameworks create demand for data that traditional methods cannot fully satisfy, generating political will and financial resources for data innovation.

Tanzania's commitment to these global frameworks provides justification for investments in Big Data capabilities. By framing Big Data adoption as essential for SDG monitoring and reporting, advocates can secure high-level support and resources that might otherwise be unavailable.

ii. International Coordinating Agencies

The United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) has developed a global strategic framework on geospatial information and services for disasters whose objective is to help countries benefit from using geospatial information and services across the entire emergency cycle. Although mainly focused on

humanitarian applications, such initiatives can significantly benefit availability and accessibility of quality geospatial datasets for health, environment, and other sectors.

These international coordinating bodies provide technical guidance, facilitate knowledge exchange, and sometimes offer direct support to member countries. Tanzania can leverage these relationships to access expertise, methodologies, and technologies that would be difficult to develop independently.

iii. Data-focused collaboratives

Initiatives such as the Health Data Collaborative (HDC) and the Global Partnership for Sustainable Development Data support countries in improving availability, quality, and use of data for local decision-making by keeping data high on the political agenda, aligning efforts to improve demand and supply of data at national levels, and improving data use through development of reference standards, norms, and practical tools.

For instance, the Open Health Information Exchange (OpenHIE) initiative develops interoperable, standards-based health information exchange systems for resource-constrained settings. These collaborative platforms provide Tanzania with ready-made solutions and communities of practice that can accelerate Big Data adoption while avoiding common pitfalls.

iv. Regional knowledge hubs

Increasing investment by countries and donor agencies in establishing regional knowledge hubs supports countries in strengthening use of Big Data and new technologies. Examples include the WHO/AFRO Regional GIS Center for the Polio Eradication Program established with support from the Bill and Melinda Gates Foundation and the multi-sectoral Regional Centre for Mapping of Resources for Development (RCMRD) established in Kenya.

Access to available resources at regional level can be achieved by promoting a culture of networking and collaboration in Tanzania at national, East African Community (EAC), and global levels. Regional hubs offer advantages of cultural and contextual relevance while providing economies of scale in capacity building and infrastructure development.

v. Partnership with Data Holders at National Level

Without access to the necessary datasets, the data innovation projects will never advance. Engagement in an active data partnership at national level is essential in advancing and advocating the application of big data in Tanzania. This can be achieved by formal collaboration (agreements or legislation) between governments, private sector, researchers, data providers, academia and tech companies to unlock access to non-traditional data and to get empowered through knowledge and skills transfer, as well as infrastructures.

The stakeholder survey revealed strong interest in partnerships, with 42.9 percent planning international partnerships and 36.4percent planning national partnerships. This indicates readiness for collaborative approaches. The challenge lies in creating enabling frameworks and mechanisms that facilitate these partnerships while protecting legitimate interests of all parties.

Successful partnerships require clear value propositions for all participants. Private companies need assurance that data sharing will not compromise competitive advantages or customer privacy. Government agencies need confidence that partners will use data responsibly and ethically. Academic institutions need guarantees of research freedom and attribution. Well-designed partnership frameworks can balance these interests while maximizing collective benefits.

vi. Strengthening National Data Infrastructures

At the national level, awareness and efforts are growing to create institutional frameworks facilitating production, standardization, and sharing of geospatial and Big Data crucial across a wide range of sectors. Tanzania can leverage already existing National Data Infrastructures such as Tanzania Data Lab (dLab), Tanzania Integrated Statistical Portal (TISP), and University-Affiliated Hubs.

Such infrastructure can maximize use and minimize redundant creation of geospatial information for use across sectors, including for social and economic development. To maximize Big Data impact, collaboration among governments, private sectors, academic institutions, and international organizations remains essential.

These existing platforms provide foundations upon which to build. Rather than creating entirely new systems, Tanzania can enhance and expand current infrastructure, adding Big Data

capabilities incrementally. This approach reduces costs, builds on proven systems, and creates continuity with existing workflows and user bases.

vii. Availability and Accessibility of Big Data

From a data availability perspective, an increasing number of global and regional datasets are accessible for public use. While these datasets might not be validated by countries, they represent useful sources when official country data are not available or accessible. Among these datasets are Google Earth Engine (GEE), OpenStreetMap (road network, hydrographic network, populated places, etc.), WorldPop, FAO databases, Demographic and Health Survey (DHS, for demographic, health, and development indicators), OpenAerialMap, GlobeLand30 and the Global Land Cover Facility (land cover, satellite images, and other remote sensing products), Humanitarian Data Exchange (HDX), and the Group of Earth Observations System of Systems (GEOSS), among others.

These global datasets offer Tanzania several advantages. First, they provide immediate access to data that would be costly and time-consuming to generate domestically. Second, they offer opportunities for benchmarking and validation of national data. Third, they can fill gaps in spatial or temporal coverage where national data collection is limited. Fourth, they enable Tanzania to participate in global monitoring initiatives and comparative analyses.

However, reliance on global datasets also presents challenges. Data may not align perfectly with national definitions or classifications, requiring careful harmonization. Global datasets may lack the granularity needed for subnational planning and monitoring. Quality and accuracy may vary across geographic areas, with some regions better represented than others. Therefore, these global resources should complement rather than replace national data collection efforts.

viii. Geospatial Software and Tools

It is worth noting that access to geospatial technologies is no longer a significant barrier due to competitive geospatial software and services markets. The increasing availability of low-cost or free and open-source software and tools can facilitate access to and use of Big Data analytics. Open-source platforms such as QGIS, R, Python with geospatial libraries, and cloud-based platforms like Google Earth Engine provide powerful capabilities without licensing costs.

This democratization of technology means that financial constraints need not prevent Big Data adoption. Small organizations, academic institutions, and government agencies with limited budgets can access world-class analytical tools. The challenge shifts from acquiring software to building capacity to use it effectively and maintaining the necessary hardware and network infrastructure.

Open-source tools also offer additional benefits including transparency in analytical methods, active user communities providing support and resources, flexibility to customize solutions for specific needs, and independence from vendor lock-in. These advantages are particularly relevant for Tanzania as it builds sustainable Big Data capabilities.

ix. Investment in Human Capacity Building

Strategic investment in human capacity building aims at producing strong data scientists capable of understanding and using analytical tools, methods, and new technologies necessary to work with Big Data and extract insights from datasets. Equally important are data engineers and architects capable of establishing architecture and technical infrastructure needed to transfer, store, and process Big Data efficiently and securely.

Tailored training programs should focus on data science, machine learning, geospatial analysis, and data privacy to ensure that staff can handle emerging data sources like satellite imagery, mobile phone records, social media, and citizen-generated data. However, capacity building must extend beyond technical skills to include:

Statistical Methodology: Understanding how to adapt Big Data for statistical production while maintaining quality standards and addressing bias and representativeness issues.

Domain Knowledge: Combining technical skills with deep understanding of specific sectors such as agriculture, health, or environment to ensure analyses are contextually appropriate and policy-relevant.

Data Ethics and Governance: Training on responsible data use, privacy protection, informed consent, and ethical considerations in Big Data applications.

Communication Skills: Ability to translate complex analytical results into actionable insights for policymakers and communicate findings to diverse audiences.

Project Management: Skills to plan, implement, and sustain Big Data initiatives within organizational contexts.

Capacity building should employ diverse approaches including formal academic programs, short-term specialized training, on-the-job learning, mentorship programs, communities of practice, and exchange programs with institutions that have advanced Big Data capabilities. Importantly, capacity building must be continuous rather than one-off, recognizing the rapidly evolving nature of Big Data technologies and methods.

x. Strengthen and Establish a Space for Data Innovation

Keeping data and emerging technologies high on the political agenda and encouraging investments in data innovation across various fields requires dedicated institutional spaces. Strengthening existing data innovation centers such as COSTECH, establishing and operationalizing new data science campuses, centers of excellence, new Big Data and data innovation units in organizational structures, and innovation hubs creates spaces for Big Data skills development and expertise.

These innovation spaces serve multiple functions. They provide physical and institutional environments where experimentation with Big Data is encouraged and supported. They create communities of practice bringing together diverse stakeholders including technologists, statisticians, domain experts, and policymakers. They serve as demonstration sites showcasing successful Big Data applications that can inspire broader adoption. They function as training grounds where practitioners gain hands-on experience with real-world data and problems.

Innovation hubs can also play crucial roles in bridging gaps between research and practice, academia and government, and national and international expertise. By creating neutral spaces where different actors can collaborate, innovation hubs facilitate the partnerships and knowledge exchange essential for Big Data ecosystem development.

For Tanzania, investing in such spaces represents strategic allocation of resources. Rather than dispersing limited funds across many small initiatives, concentrating investment in well-designed innovation hubs can create critical mass of expertise and infrastructure that benefits the entire ecosystem. These hubs can also serve as focal points for international partnerships and donor support.

3.2 Conclusion

Tanzania stands at a critical juncture in its data development journey. The convergence of global commitments under the Sustainable Development Goals, rapid digital transformation, growing recognition of data's value for development, and increasing availability of Big Data sources creates unprecedented opportunities for strengthening evidence-based policymaking and development planning.

This assessment has documented both the challenges Tanzania faces in Big Data adoption and the significant opportunities available. The stakeholder survey revealed that while Big Data use remains limited and significant capacity gaps exist, there is strong awareness of potential benefits and considerable willingness to pursue data innovation. International case studies demonstrate that countries at similar development stages have successfully implemented Big Data solutions generating tangible benefits for policy and development.

The three showcase areas examined; environment, agriculture, and poverty statistics illustrate broader patterns applicable across sectors. Traditional data collection methods, while valuable, cannot alone meet the growing demand for timely, granular, and comprehensive data required for effective development planning and monitoring. Big Data and new technologies offer complementary approaches that can enhance rather than replace traditional statistics, providing more frequent updates, greater spatial granularity, and insights into previously unmeasurable phenomena.

However, realizing Big Data's potential requires addressing fundamental challenges in institutional frameworks, human capacity, data quality and accessibility, and awareness. These challenges are interconnected and mutually reinforcing solving them requires comprehensive, coordinated approaches rather than isolated interventions.

The opportunities identified from global frameworks and partnerships to data infrastructure and innovation spaces provide practical pathways forward. Tanzania need not develop everything from scratch; substantial resources, expertise, and technologies are available through international partnerships, regional collaborations, and global platforms. The key is strategic engagement with these resources, adapting international experience to Tanzania's specific context while building sustainable local capacity.

Moving forward, Tanzania should pursue a balanced approach combining quick wins from pilot projects demonstrating value with longer-term investments in capacity, infrastructure, and institutional frameworks ensuring sustainability. Strategic priorities include developing comprehensive policy frameworks for Big Data governance and data sharing, making targeted infrastructure investments leveraging existing platforms, implementing multi-faceted capacity building strategies addressing immediate needs while building long-term expertise, pursuing strategic pilot projects in priority areas, systematically developing partnerships at national, regional, and international levels, and conducting awareness and advocacy campaigns building demand and political support.

Success will require sustained commitment from government leadership, active engagement from diverse stakeholders, adequate resource allocation, and patience recognizing that building robust Big Data ecosystems takes time. However, the potential rewards improved policy effectiveness, enhanced development outcomes, better service delivery, and progress toward national and global development goals justify the required investments.

The Tanzania Statistical Master Plan II (TSMP II) provides an operational roadmap, and this assessment report serves as the overarching, integrative tool empowering and expanding upon TSMP II objectives. By moving beyond production of statistics using traditional data sources to accommodate and complement non-traditional data sources and actors from private sector and civil society, Tanzania can create a unified, ethical, and innovative data ecosystem supporting evidence-based development.

Tanzania's journey toward effective Big Data integration is challenging, but it is necessary and achievable. With strategic vision, collaborative spirit, sustained commitment, and willingness to learn from both successes and failures, Tanzania can build the data capabilities required to achieve its development aspirations and fulfil its commitments to leaving no one behind in the pursuit of sustainable development. The time to act is now, and this assessment provides the foundation for transforming vision into reality.

REFERENCES

Australian Bureau of Statistics, Sugarcane, experimental regional estimates using new method data sources and methods, 17 June 2020, https:www.abs.gov.au/statistics/industry/agriculture/sugarcane-experimental-regional-estimates-using-new-data-sources-and-methods/2019-20

Best practices in innovations in microplanning for polio eradication: supplement to best practices in microplanning for polio eradication. Geneva: World Health Organization; 2018(WHO)/POLIO/18.08). Licence: CC BY-NC-SA 3.0 IGO.

Chen, S., Chen, B., & Gong, P. (2016). Big data for environmental sustainability: Progress and prospects. *Annual Review of Environment and Resources, 41*, 599–622.

De Mauro, A., Greco, M., Grimaldi, M., 2016. A formal definition of Big Data based on its essential features. Libr. Rev. 65, 122–135.

FAO, IFAD and WFP (2015), The State of Food Insecurity in the World 2015. Meeting the 2015 international hunger targets: taking stock of uneven progress, Food and Agriculture Organization Publications, Rome.

Faulkner, A., Cebul, K., 2014. Agriculture Gets Smart: The Rise of Data and Robotics, Cleantech Agriculture Report. Cleantech Group.

Fenn, J., LeHong, H., 2011. Hype Cycle for Emerging Technologies, 2011. Gartner (July).

Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. Journal of Business Research, 108, 147–162. doi: 10.1016/j.jbusres.2019.09.062.

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Ullah Khan, S., 2015. The rise of "Big Data" on cloud computing: Review and open research issues. Inf. Syst. 47, 98–115.

Li, X., Zhou, Y., Chen, Y., & Chen, Z. (2017). Using big data to predict and assess the impact of urban expansion on air quality. *Environmental Pollution, 220*(Pt B), 1322–1330.

Liu, J., Ye, Z., Tan, R., & Ma, L. (2018). Modeling the environmental and economic impacts of coal-to-renewable energy transition in China using big data. *Applied Energy, 225*, 182–194.

Mikalef, P., & Krogstie, J. (2020). Examining the interplay between big data analytics and contextual factors in driving process innovation capabilities. European Journal of Information Systems, 29(3), 260–287. doi: 10.1080/0960085x.2020.1740618.

Mohapatra, S. K., & Mohanty, M. N. (2020). Big data analysis and classification of biomedical signal using random forest algorithm. New Paradigm in Decision Science and Management: Proceedings of ICDSM 2018 (pp. 217–224). doi: 10.1007/978-981-13-9330-3_20.

Needle, D., 2015. Big Data or Big Disappointment? Experts Debate Hype Versus Reality. eWeek. http://www.eweek.com/database/big-data-or-big-disappointment-experts debate-hype-versus-reality.html (Accessed: 2 August 2016).

Russom, P. (2011). Big data analytics. TDWI Best Practices Report, Fourth Quarter, 19(4), 1–34.

Semantic Community, 2015. Big Data Science for Precision Farming Business. http://semanticommunity.info/Data_Science/Big_Data_Science_for_Precision_Farming_ Business (Accessed: 2 August 2016).

Shukla, A. K., Muhuri, P. K., & Abraham, A. (2020). A bibliometric analysis and cutting-edge overview on fuzzy techniques in Big Data. Engineering Applications of Artificial Intelligence, 92, 103625. doi: 10.1016/j.engappai.2020.103625.

Sun, H., Chen, Y., Li, J., & Liu, Q. (2021). Big data analytics for assessing the environmental impacts of industrial activities in coastal regions. *Journal of Big Data, 8*(1), 1–18.

Sundmaeker, H., Verdouw, C., Wolfert, S., Pérez Freire, L., 2016. Internet of food and farm 2020. In: Vermesan, O., Friess, P. (Eds.), Digitizing the Industry- Internet of Things Connecting Physical, Digital and Virtual Worlds. River Publishers, Gistrup/Delft, pp. 129–151.

Tran, L. Q. T. (2022). Building a conceptual framework for using big data analytics in the banking sector. Intellectual Economics, 1(16).

Wang, J., Zhao, S., Li, W., & Liu, J. (2018). Integrating big data into life cycle assessment for water footprint analysis of agricultural products. *Journal of Cleaner Production, 172*, 2340–2351.

Wolfert, J., Sorensen, C.G., Goense, D., 2014. A Future Internet Collaboration Platform for Safe and Healthy Food from Farm to Fork, Global Conference (SRII), 2014 Annual SRII. IEEE, San Jose, CA, USA, pp. 266–273.

Zhang, Y., & Wang, X. (2023). Optimizing energy consumption in manufacturing processes using a hybrid big data modeling approach. *Energy, 238*, 121687.

Zhou, X., Chen, S., & Zhang, Y. (2023). A framework for assessing the environmental impacts of transportation infrastructure projects using multi-source big data. *Environmental Impact Assessment Review, 101*, 106664.

Appendix A: List of Selected SDG targets that could be supported with Big Data Sources

SDG Target	Big Data source for the SDG target	SDG Indicator	Big data needed or could support	Relevance to other statistical frameworks
Goal 1: End poverty in all of its forms everywhere				
1.1 By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day	**Mobile phone data** for socioeconomic status and wellbeing: • Human mobility and socioeconomic levels • Estimating poverty and wealth • Socioeconomic status **Satellite data** for poverty mapping (nighttime lights): • Identifying the poor urban poverty	1.1.1 Proportion of the population living below the international poverty line by sex, age, employment status and geographic location (urban/rural)	• **Geospatial data** can support compilation for this indicator; • **Citizen-generated data** could provide supplementary information	
1.2 By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	**Mobile phone data** for socioeconomic status and wellbeing: • Human mobility and socioeconomic levels • estimating poverty and wealth	1.2.1 Proportion of population living below the national poverty line, by sex and age	• **Citizen-generated data** could provide supplementary information	

SDG Target	Big Data source for the SDG target	SDG Indicator	Big data needed or could support	Relevance to other statistical frameworks
	• socioeconomic status **Satellite data** for poverty mapping: • identifying the poor urban poverty			
		1.2.2 Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	• **Citizen-generated data** can have a direct contribution	
1.4 By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance	**Mobile phone data** for financial inclusion: creditworthiness of the unbanked (airtime credit data)	1.4.1 Proportion of population living in households with access to basic services	• **Citizen-generated data** could provide supplementary information	
		1.4.2 Proportion of total adult population with secure tenure rights to land, (a) with legally recognized documentation, and (b) who perceive	• **Geospatial data** can support compilation for this indicator (UN-GGIM/GEO); • **Citizen-generated data** can have a direct	

SDG Target	Big Data source for the SDG target	SDG Indicator	Big data needed or could support	Relevance to other statistical frameworks
		their rights to land as secure, by sex and type of tenure	contribution	
1.5 By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters	Mobile phone data for disaster response: Human mobility after disasters	1.5.1 Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population	• **Geospatial data** can support compilation of this indicator (GEO); • **Mobile data** can contribute indirectly towards the measurement of this indicator; • **Citizen-generated data could provide supplementary information** or can have a direct contribution	*Sendai Framework Indicators A1 and B1*
		1.a.2 Proportion of total government spending on essential services (education, health and social protection)		
Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture				
2.1 By 2030, end hunger and ensure access by all people, in particular the poor and people in vulnerable situations, including infants, to safe, nutritious and sufficient food all year round	**Mobile phone data** and social media data for expenditure on food: • proxy indicator for food expenditure	2.1.1 Prevalence of undernourishment		

SDG Target	Big Data source for the SDG target	SDG Indicator	Big data needed or could support	Relevance to other statistical frameworks
	Satellite data for drought monitoring severity and extent of drought conditions			
2.4 By 2030, ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, that help maintain ecosystems, that strengthen capacity for adaptation to climate change, extreme weather, drought, flooding and other disasters and that progressively improve land and soil quality	• **Satellite data** for early crop yield assessment; • Developing vegetation health indices	2.4.1 Proportion of agricultural area under productive and sustainable agriculture	• **Geospatial data** is needed to compile the indicator; • **Citizen-generated data** can have a direct contribution	
13.1 Strengthen resilience and adaptive capacity to climate- related hazards and natural disasters in all countries	• Mobile phone data can support estimation of human mobility after disasters	13.1.1 Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population	• **Geospatial data** can support compilation for this indicator; • **Mobile phone data** can contribute towards measurement	*Sendai Framework Indicators A1 and B1*
13.3 Improve education, awareness-raising and human and institutional capacity on	• Satellite data can inform changes in the water- related ecosystem and	13.3.1 Extent to which (i) global citizenship education and (ii) education for sustainable		

SDG Target	Big Data source for the SDG target	SDG Indicator	Big data needed or could support	Relevance to other statistical frameworks
climate change mitigation, adaptation, impact reduction and early warning	drought monitoring	development are mainstreamed in (a) national education policies; (b) curricula; (c) teacher education; and (d) student assessment		
15.1 By 2020, ensure the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular forests, wetlands, mountains and drylands, in line with obligations under international agreements	• **Satellite data** can be used for forest mapping	15.1.1 Forest area as a proportion of total land area	**Geospatial data** is needed to compile this indicator	*Ecosystem Extent / Land Cover Account*
15.3 By 2030, combat desertification, restore degraded land and soil, including land affected by desertification, drought and floods, and strive to achieve a land degradation- neutral world	• Satellite data can be used to assess changes in vegetation	15.3.1 Proportion of land that is degraded over total land area	• **Geospatial data** is needed to compile this indicator	*Ecosystem Condition Account and Ecosystem Extent / Land Cover Account*