New Data Eco-System and Official Statistics in Africa – Maximizing the Opportunities

Instructions: Click on the link to access each author's presentation.

Organiser: Osuolale Peter Popoola

Chair: Misha Belkindas

Participants:

Osuolale Peter Popoola: Importance of Data in Africa Development

Oliver J. M Chinganya: Building public trust in the use of big data for official statistics

Samuel K. Annim:* Big data and the modernization of statistical Systems in Africa

Prince Semiu Adeyemi Adeniran:* Harnessing the full power of Timely Data for Official Statistics in Africa

*Work presentation not available or non-existent

Table of Content

➢ Concept of development

➢ Importance of Data to Development

➢ Situation of Data in Africa

➢ The New Data Eco-System: Opportunity for Africa

➢ The Big Data Revolution

➢ NSOs in Africa and Big Data Project

➢ Conclusion

Concept of Development

- Development refers to the entire spectrum of aspects that advance the quality of human life.
- These include health, education, trade and employment, equality, environment and natural resource allocation and sustainability, social and political stability
- Its summed up all economic and non-economic spheres i.e social-economic development.
- WB defines 'development data' as the data produced by country systems (e.g. NSOs, government line ministries and agencies) or other parties on a country's social, economic and environmental issues.
- Which includes data sourced from census, household surveys, agricultural surveys, civil registration and vital statistics and administrative systems.
- The term also encompasses additional forms of data (big, open, citizen-generated and real-time digital data).
- There are different sources and types of data which often take various formats and which could be used to inform development interventions.

- Development data are also compiled by international organizations and financial institutions to monitor the pace of economic and social development as well as status of the environment.

- It provide essential insights to solving human problem;

- It helps in understanding the practicalities of development process;
- interactions and feedback among different systems;
- Is the factors that shape decisions.
- Development is held back when economic and demographic data that form the basis for decision making are lacking or insufficient to form a complete picture of what is needed.

- A country should be able to provide a quality, adequate, reliable and accessible data to access its growth and monitor development.
- However, producing more data is not enough:
- Data must be transformed, analyzed and useful for:
- policy making;
- monitoring and;
- accountability.
- When data produced could not meet the above, it means development data is not available.

Data are a prerequisite for measuring development indicators.
- Sources of data for computation of development indicators include:
➢ Censuses;
➢ Sectoral surveys;
➢ Administrative; and
➢ Civil registration or vital statistics

Development data are also refers to Official Statistics.

Importance of data to Development

- Data are crucial for effective policymaking.
- Data helps in delivery of public goods and services
- Data is useful to democracy

- Data are the best steering wheel for policy; a tool with which to govern
- Data is like a dashboard to a car.
- No data, no development

Situation of Data in Africa

❖ **Big question: Is there Adequate Data for** essential development indicators in Africa?

❖ Many **developing nations** do have

✓ inadequate poverty data: income, expenditure, e.t.c
✓ Incomplete civil registration and vital statistics
✓ Lack of disaggregated data – by gender and disability
✓ Lack of effective legal framework
✓ Lack of financing/fund for quality data collection.

Data Type	Tools	Africa Status
Civil registration	Vital statistics, censuses, Household Surveys	➢ Less than one in five known births occurs in a country with a complete birth registration system. ➢ In Africa, 87% of deaths occur in countries without a complete death registration system
Population	Censuses	➢ Almost nine out of ten people live in a country which has conducted a population census in the last ten years. ➢ A third of all Africans live in a country where a census has been conducted since 2010
Poverty & inequality	Household Surveys	➢ Almost all (99%) Africans live in a country which has conducted a Household Survey in the last ten years. ➢ Despite this, only half of the continent's population lives in a country that has carried out more than two comparable Household Surveys in the past ten years. ➢ For half the continent's population, changes in levels of poverty are unknown
Economic growth	National accounts, administrative data	➢ Only seven countries in Africa use the 2008 UN System of National Accounts, the latest version of the international statistical standard for measuring macroeconomic indicators. ➢ Less than a third of countries in Africa have produced industrial data since 2006. However, 45 countries in Africa have produced trade statistics in the last ten years

Employment	Labour Force Surveys	➢ Over half of African citizens live in a country which has not conducted a Labour Force Survey in the past ten years. ➢ This means key indicators around the labour market and employment are unknown in these countries
Agriculture	Agricultural census	➢ Just over half of Africans live in a country which has conducted an agricultural census in the last ten years. ➢ For almost half of the continent's population, information around structure of the agricultural sector and landholders is unknown
Safety	Administrative data	➢ Only ten countries recorded the prevalence of drug usage in the United Nations Office on Drugs and Crime Homicide Statistics database in the last ten years
Health	Administrative data	➢ Since 2005, 80% of countries published a Household Survey including a health component
Education	Administrative data	➢ Only 29% of countries have published a Household Survey including an education component since 2005.

The New Data Ecosystem: Opportunity for Africa

Our world is changing rapidly(Statistics)

The earliest data collections took the form of census.

Sampling was discovered and gave rise to surveys and then multi-topic surveys

The Digital revolution of Big data is changing our world

One of the Characteristics of the 4th Industrial revolution is the internet of things.

Before this era, official data are generated largely from sample survey, censuses and administrative methods

❏ At its core, data represents a post-industrial opportunity.
- The world create 2.5 exabytes [10^18] bytes of data every day
- 90% of the data in the world created in the last two years through the use of internet, social media, commercial transactions, digital images etc.
- In 2020, the digital universe was estimated to consist of 44 zeta-bytes of data
- Predicted that by 2025,
- 175 Zeta-bytes data would be created every 24 hours worldwide
- Big Data is a collection of data sets so:
  ➢ large and complex
  ➢ becomes difficult to process using on-hand database management tools;
  ➢ or traditional data processing applications

- Big data burst upon the scene in the first decade of the 21st century

- Like many new information technologies, big data can bring about dramatic cost reductions, substantial Improvements in obtaining official statistics by the NSOs in Africa.

- From the dawn of civilization until 2013, humankind generated 5 Exabyte's of data. Now the world is producing 5 Exabyte's of data every two days...and the pace is accelerating

What is Big Data
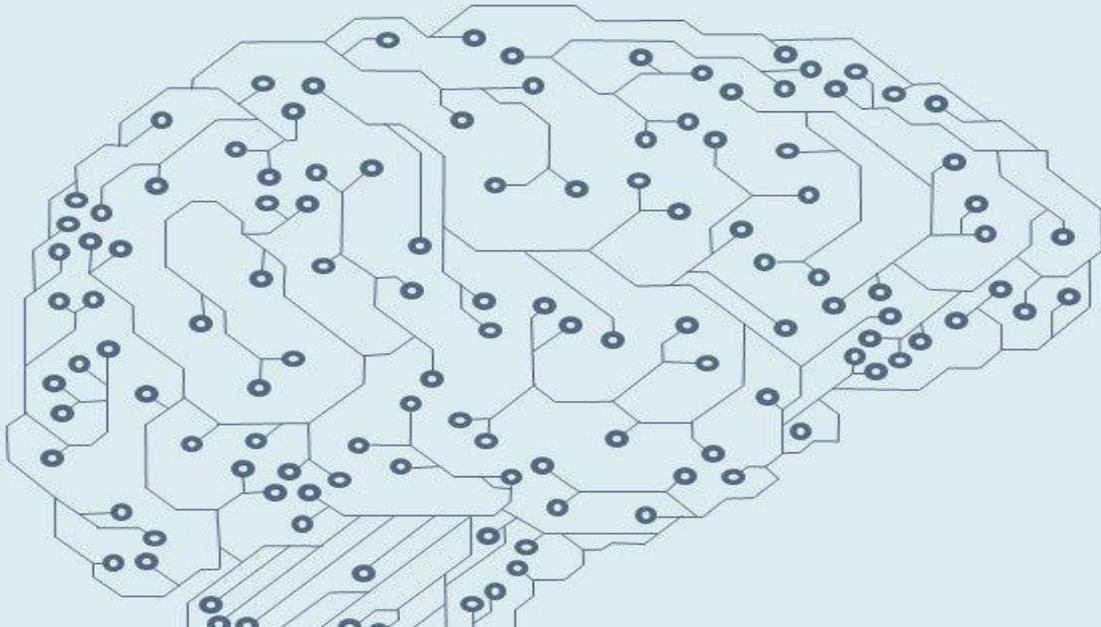
Big Data

Big data is characterized by large volumes of different types of data (e.g. Social, web, transaction, etc.) That builds very quickly.

It exceeds the reach of commonly used hardware environments and software tools to capture, manage and process in a timely manner for its users.

Big Data Revolution
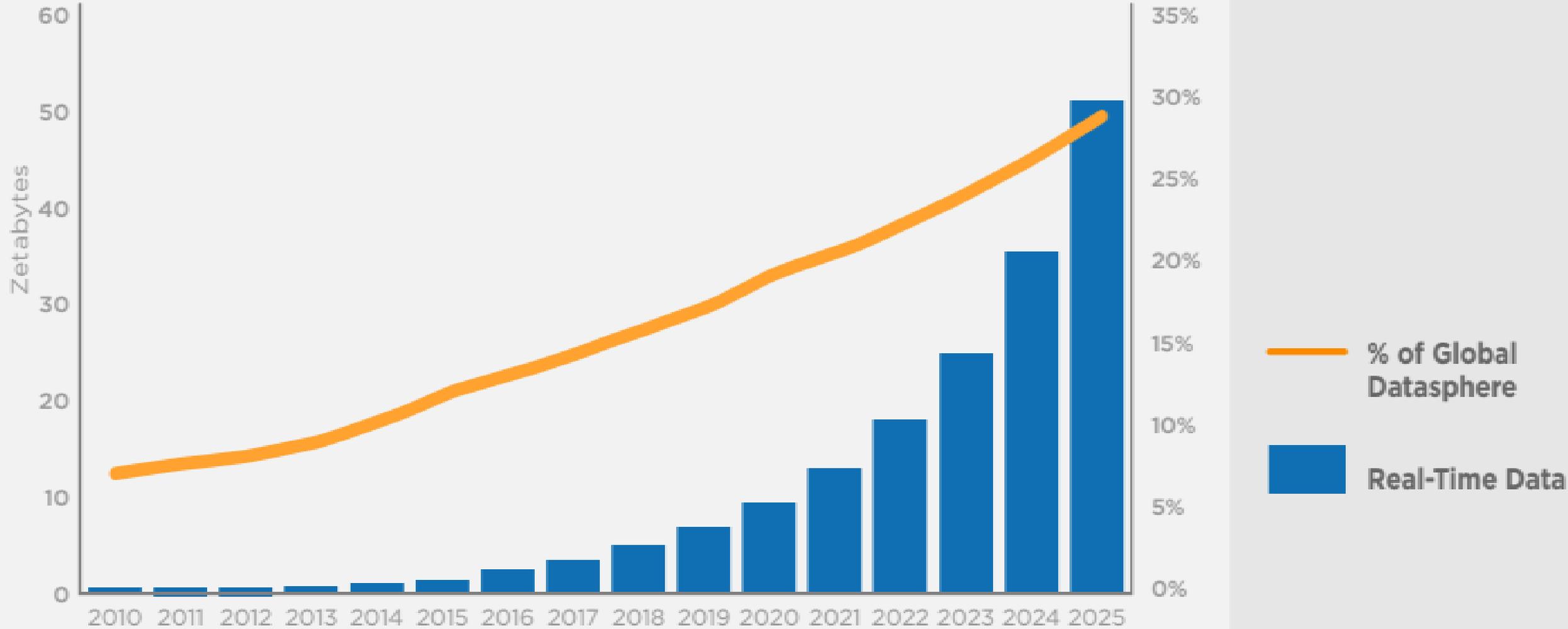
Figure 1 – Annual Size of the Global Datasphere

Annual Size of the Global Datasphere

175 ZB
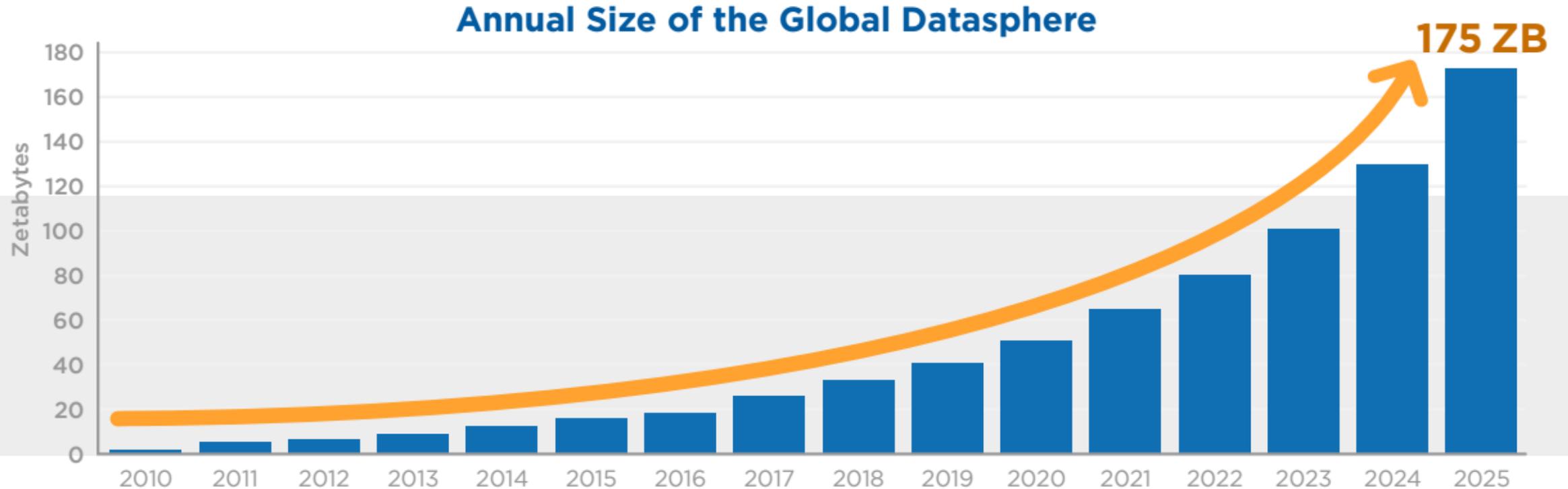
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Figure 8 - Real-Time Data

How Much of Global Datasphere is Real-Time?

Legend:
- % of Global Datasphere (orange line)
- Real-Time Data (blue bars)

- As of June 2022, there were more than 4.5 billion people online;
- 80% of digital content is available in nine out of every ten languages
- Google processed 3.7 million queries;
- More than 1 million logins on Facebook Daily;
- YouTube recorded 4.5 million videos viewed every 60 seconds;
- In the Social Media the data generated every **minute** include Snapchat users share 527,760 photos;
- more than 120 professionals join LinkedIn; 4,146,600 YouTube videos; 456,000 tweets; and Instagram users post 46,740 photos.
- More than a quarter of the world's 7 billion humans, over 2 billion are active on Facebook.
- Facebook alone have 1.5 billion people are active daily;
- 5 new Facebook profiles created every second; and
- More than 300 million photos get uploaded per day

NSOs in Africa and Big Data Project

In Africa

Europe and Developed countries

Currently in Africa

Most NSOs and Research Institutes are still facing the challenges of obtaining official statistics using traditional methods, some handful countries like Kenya, Zambia had experiment the use of mobile phone for their PHC

NSOs and some agencies are using a variety of sources, including satellite imagery, aerial imagery, mobile phone data, among others to compile a wide range of official statistical indicators e.g Eurostat, UNSD and UNE

more than half of the SDGs indicators have no or insufficient data for its monitoring

NSOs and research institutes in Africa need to incorporate big data in obtaining OS

- ➢ The rise in big data should change the context in which (NSOs) in Africa operate.
- ➢ Given big data scope,
- ➢ timeliness,
- ➢ resolution,
- ➢ and the potential efficiencies it offers in obtaining official statistics
- ❑ As the world moves into the 4th Industrial era, more and more data are generated on the web, social media, mobile phones, and censors in the ever-growing number of electronic devices around us.
- ➢ Big Data provides opportunities to obtain
- ➢ timely,
- ➢ costless,
- ➢ higher precision,
- ➢ completeness, and
- ➢ less burden official statistics.

- Big data can also be used to monitor the 17 Sustainable Development Goals (SDG) and
- its 169 indicators, especially where traditional data collections are missing.
- Currently in Africa, more than half of the SDG indicators have no or insufficient data to measure progress, big data has the potential to address some important gaps
- For Africa development, NSOs in Africa need to leverage in the use of the New Data Eco-system
- In years to come, developments such as Data Science, Artificial Intelligent, Machine Learning, Cloud Computing will surely change the face of official statistics.
- Big Data may offer new cost-effective or timely ways of compiling official statistics or offer some relief to survey fatigue and burden.
- The Big Data potentials are generally accepted within official statistics circles for enhanced insight and decision making.
- These potentials may be used in conjunction with or as a replacement for traditional data sources to improve, enhance and complement existing official statistics.

Thank you

IOS-ISI 2024
MEXICO CONFERENCE

Building Public Trust in Big Data and Machine Learning for Official Statistics in Africa

INEGI   isi International Statistical Institute   IAOS IMPROVING OFFICIAL STATISTICS

Oliver Chinganya,

Director, African Centre for Statistics

Chief Statistician, United Nations Economic Commission for Africa

Building Public Trust in Big Data and Machine Learning for Official Statistics in Africa

Oliver Chinganya, Director, African Centre for Statistics

United Nations Economic Commission for Africa

IAOS-ISI 2024, Mexico City

Improving Decision-Making for All

New Data Eco-System and Official Statistics in Africa – Maximizing the-Opportunities

May 16 2024

Introduction

- **Overview**
  - **Revolutionizing Statistics:** Big data and machine learning are transforming the capabilities for data analysis.
  - **Enhanced Decision Making:** Enabling faster, more accurate insights for policy and resource allocation.

- **Context**
  - **Tailored for Africa:** Addressing unique challenges in agriculture, health, and economic planning.
  - **Supporting Development:** Critical tools for achieving Sustainable Development Goals and responsive policies.

- **Objective**
  - **Building Trust:** Ensuring the ethical use and transparency of big data and machine learning.
  - **Community Engagement:** Involving stakeholders in understanding the benefits and risks.

Opportunities of Big Data in Official Statistics

- **Enhancing Statistical Capabilities**
  - **Instantaneous Statistics:** Utilize real-time data from mobile and social media for faster response and analysis.
  - **Dynamic Insights:** Ability to adapt to rapid changes and provide up-to-date information.

- **Improving Accuracy and Timeliness**
  - **High-Volume Data Analysis:** Harness the vast amounts of data for deeper insights and precision.
  - **Predictive Analytics:** Use machine learning to forecast trends and potential impacts on society.

- **Enabling Informed Decision Making**
  - **Policy Impact:** Data-driven insights to support effective policy making.
  - **Resource Allocation:** Optimize distribution of resources based on accurate, real-time data.

Challenges in Integrating Big Data

- **Methodological Issues**
  - **Data Volatility:** Addressing the inconsistency and unpredictability of big data sources.
  - **Integration Complexity:** Combining big data with traditional data sources to ensure accuracy.
- **Data Quality and Reliability**
  - **Quality Assurance:** Developing new methods to verify the quality of data derived from diverse sources.
  - **Reliability Concerns:** Ensuring that statistics produced are consistent and dependable over time.
- **Addressing Privacy and Security**
  - **Data Protection:** Implementing robust security measures to protect data integrity and confidentiality.
  - **Privacy Concerns:** Navigating the complexities of personal data usage while respecting privacy laws.
- **Ethical and Legal Challenges**
  - **Ethical Use:** Maintaining fairness and avoiding bias in data collection and analysis.
  - **Regulatory Compliance:** Adhering to international and local data governance standards.

Ethical Considerations and Public Trust

- **Ensuring Ethical Use of Data**
  - **Transparency:** Committing to openness in data collection, processing, and usage.
  - **Accountability:** Holding institutions responsible for the ethical use of data.

- **Building and Maintaining Public Trust**
  - **Public Engagement:** Regularly involving community and stakeholders in discussions about data use.
  - **Trust-Building Measures:** Demonstrating the benefits and safe practices of big data.

- **Privacy and Data Protection**
  - **Strong Privacy Protections:** Implementing advanced data security and anonymization techniques.
  - **Legal Compliance:** Ensuring all big data practices comply with national and international privacy laws.

- **Addressing Bias and Fairness**
  - **Bias Mitigation:** Developing techniques to detect and correct bias in data collection and analysis.
  - **Fairness in AI:** Ensuring machine learning algorithms promote equity and do not perpetuate disparities.

Technological Investment and Capacity Building

- **Infrastructure Development**
  - **Advanced IT Infrastructure:** Investing in robust hardware and software to handle large datasets.
  - **Secure Data Platforms:** Ensuring data integrity and security through state-of-the-art technologies.
- **Skill Development**
  - **Training Programs:** Implementing ongoing training for statisticians and data scientists in the latest data analysis techniques.
  - **Collaboration with Academia:** Partnering with universities to stay abreast of the latest research and innovations in big data and machine learning.
- **Capacity Building**
  - **Institutional Strengthening:** Enhancing the capabilities of National Statistical Institutes to manage and analyze big data effectively.
  - **Workforce Expansion:** Recruiting and retaining skilled personnel adept in modern statistical methodologies.
- **Innovative Technology Adoption**
  - **Embracing AI Tools:** Utilizing artificial intelligence to enhance data processing and analysis capabilities.
  - **Investment in Research and Development:** Funding projects that explore new ways of gathering, analyzing, and utilizing data.

Legal and Regulatory Frameworks

- **Developing Robust Frameworks**
  - **Creating Guidelines:** Establish comprehensive legal guidelines for big data use in statistics.
  - **Data Governance:** Implementing structured data governance models to oversee data use and ensure compliance.

- **Ensuring Compliance**
  - **Adherence to Laws:** Strict compliance with both local and international data protection laws.
  - **Regular Audits:** Conducting regular audits to ensure ongoing adherence to all regulatory requirements.

- **Ethical Standards**
  - **Setting Ethical Norms:** Developing and enforcing ethical standards specific to big data and AI in statistics.
  - **Ethical Oversight:** Establishing committees or bodies dedicated to the ethical oversight of statistical methods and practices.

- **Stakeholder Engagement**
  - **Involving Key Players:** Collaborating with government, private sector, and civil society to shape data laws.
  - **Transparency with Public:** Keeping the public informed about how data is used and protected.

Public Engagement and Transparency

- **Public Engagement Strategies**

  - **Community Outreach:** Initiating programs to educate the public on the benefits and risks associated with big data.

  - **Stakeholder Workshops:** Organizing workshops with stakeholders to discuss data use, concerns, and feedback.

- **Transparency Initiatives**

  - **Open Data Policies:** Implementing policies that allow public access to non-sensitive data and methodologies.

  - **Clear Communication:** Providing clear, understandable explanations of how data is collected, analyzed, and utilized.

- **Building Trust**

  - **Accountability Measures:** Establishing clear lines of accountability in data handling and decision-making processes.

  - **Demonstrating Benefits:** Showcasing successful case studies where big data has led to positive outcomes in public services.

- **Feedback Mechanisms**

  - **Public Consultations:** Encouraging public input on new data projects and policies to foster a sense of ownership.

  - **Responsive Adjustments:** Adapting practices based on public and stakeholder feedback to improve trust and effectiveness.

International Cooperation and Innovation

- **Global Partnerships**
  - **Collaborative Projects:** Engaging in international projects that promote the sharing of big data insights and methodologies.
  - **Standard Setting:** Participating in global forums to help establish international standards for big data and machine learning in statistics.
- **Technological Innovations**
  - **Adopting New Technologies:** Leveraging cutting-edge technologies to enhance data collection and analysis capabilities.
  - **Innovation Hubs:** Creating or participating in innovation hubs that foster the development of new statistical tools and techniques.
- **Capacity Building**
  - **Joint Training Initiatives:** Organizing international training programs to build skills in big data analytics across countries.
  - **Resource Sharing:** Facilitating the sharing of resources and expertise among national statistical offices worldwide.
- **Addressing Global Challenges**
  - **Data for Global Issues:** Utilizing big data to address transnational issues like climate change, health pandemics, and migration.
  - **Unified Responses:** Coordinating efforts to provide unified and timely data responses to global crises.

Conclusion and Way Forward

- **Summarizing Key Points**
  - **Essential Role of Trust:** Emphasizing the critical need for trust in the use of big data and machine learning.
  - **Ethical and Transparent Practices:** Reiterating the importance of ethical standards and transparency in operations.

- **Future Outlook**
  - **Continued Investment in Technology:** Highlighting the need for ongoing investment in technological infrastructure and training.
  - **Expansion of Data Capabilities:** Planning for the expansion of data collection and analysis capabilities to cover more areas.

- **Encouraging Stakeholder Engagement**
  - **Ongoing Dialogue:** Stressing the importance of maintaining an open dialogue with all stakeholders, including the public.
  - **Responsive Adaptations:** Being adaptive to feedback and changing global data landscapes to continuously improve practices.

- **Call to Action**
  - **Inclusive Participation:** Encouraging participation from various sectors in shaping the future of statistics.
  - **Commitment to Improvement:** Committing to continuous improvement and innovation in statistical methods and practices.

Thank you