



Africa Webinar, 2022

Challenges of Big Data for Official Statistics in Africa and the Way Out

Aliyu Usman, PhD.

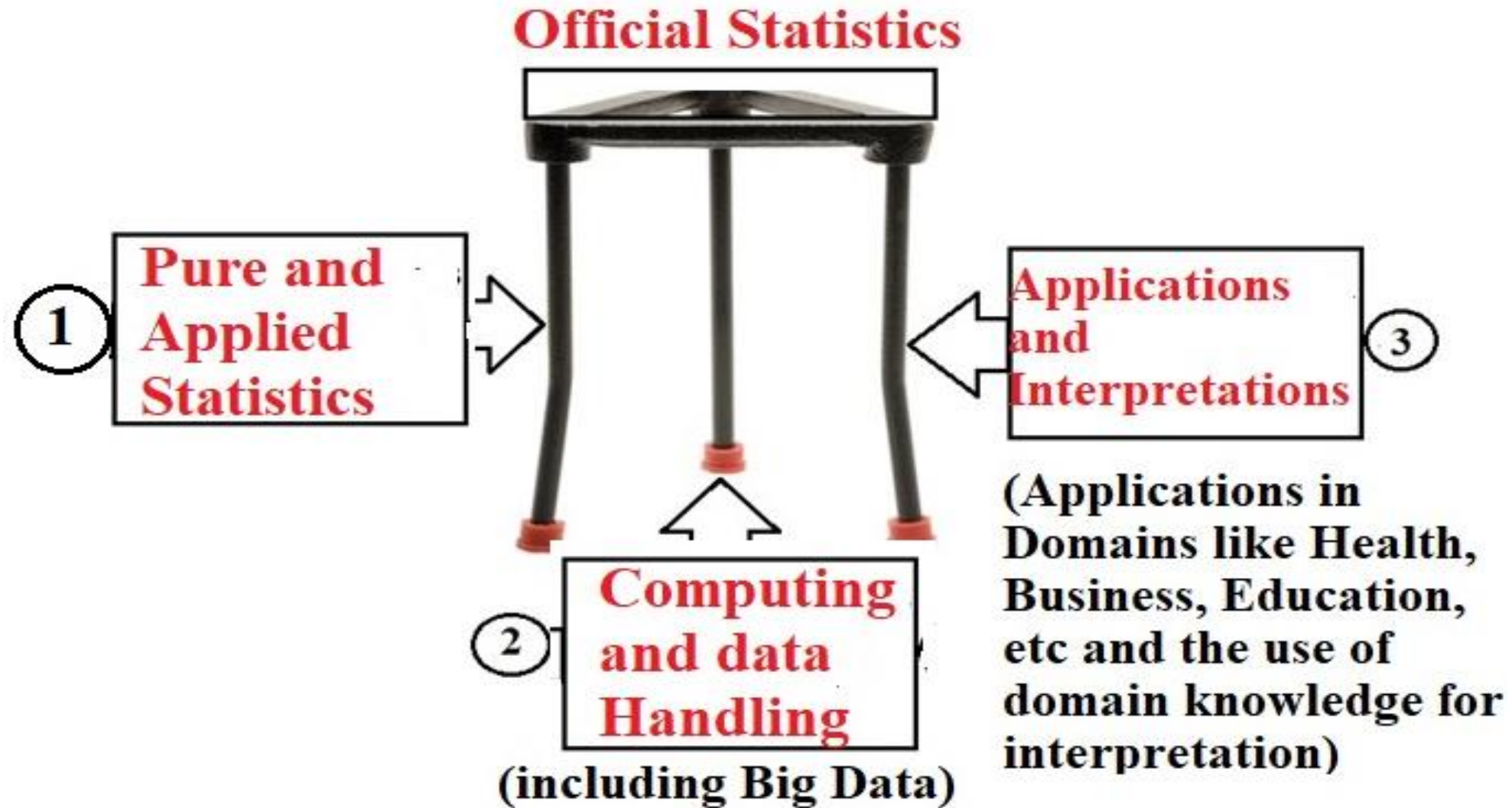
06/10/2022

Introduction

- **Official statistics are the statistical data collected, analyzed, and disseminated by governments, governmental agencies and the international bodies.**
- **Official statistics are indispensable to planning, serving the government, the economy and the public with economic, demographic, social data, among others.**
- **Big Data refers to large volume of structured, semi-structured and unstructured data that could be mined to reveal trends and relationships.**
- **Big Data influence almost all aspects of everyday life and official statistics is not left out by this influence. Hence, official statistics is a potential beneficiary of Big Data.**

Official Statistics

- **Official statistics are comprehensive data obtained from censuses or large-scale sample surveys. Censuses are usually conducted once in a decade and are supplemented by series of regular surveys that provide statistical information on annual basis, quarterly basis, or even at less frequent intervals.**
- **A chunk of official statistics also come from vital registration such as: births, deaths, marriages, crimes, notification of diseases such as cancer, HIV/AIDS, among others.**
- **Most countries conduct household surveys in-between the decennial censuses to collect economic and social data.**
- **Nigeria, for instance, do conduct the National Demographic and Health Survey (NDHS), the Multiple Indicator Cluster Survey (MICS), among others. Official statistics itself stand on three cardinal pillars. Thus:**



Fundamental Principles of Official Statistics

- **Public trust in official statistics is guided by professionalism, impartiality and use of sound methodology. In line with the Data Quality Assessment Framework (DQAF). Hence, Understanding the principles of official statistics is important.**
- **The United Nations (UN) Statistical Commission adopted the fundamental principles of official statistics in 1994 at the global level, which were then adopted by the United Nations General Assembly in 2018.**
- **The Chief Statisticians or coordinators of statistical activities of United Nations agencies and related organizations agreed that implementation of these principles will enhance the functioning of the international statistical system. The Director-General of the International Labour Organization (ILO) also became fully committed to the principles.**

Relevance, Impartiality and Equal Access

Official statistics should be compiled and disseminated on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

Professional Standards and Ethics

To gain trust in official statistics, the statistical agencies need to work strictly with professional ethics and scientific principles on the methods and procedures for the collection, processing, storage and presentation of statistical data.

Accountability and Transparency

To facilitate correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

Prevention of Misuse

The statistical agencies should clearly comment on erroneous interpretation and misuse of statistics.

Sources of Official Statistics

Data for statistical purposes may be drawn from various sources such as surveys or administrative records. Statistical agencies are to choose the source regarding quality, timeliness, costs and the burden on respondents.

Confidentiality

Data collected by statistical agencies from individuals for statistical compilation are to be treated as strictly confidential and used exclusively for statistical purposes.

Legislation

The laws and regulations under which the statistical systems operate are to be made public.

National Coordination

Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the national statistical system.

Use of International Standards

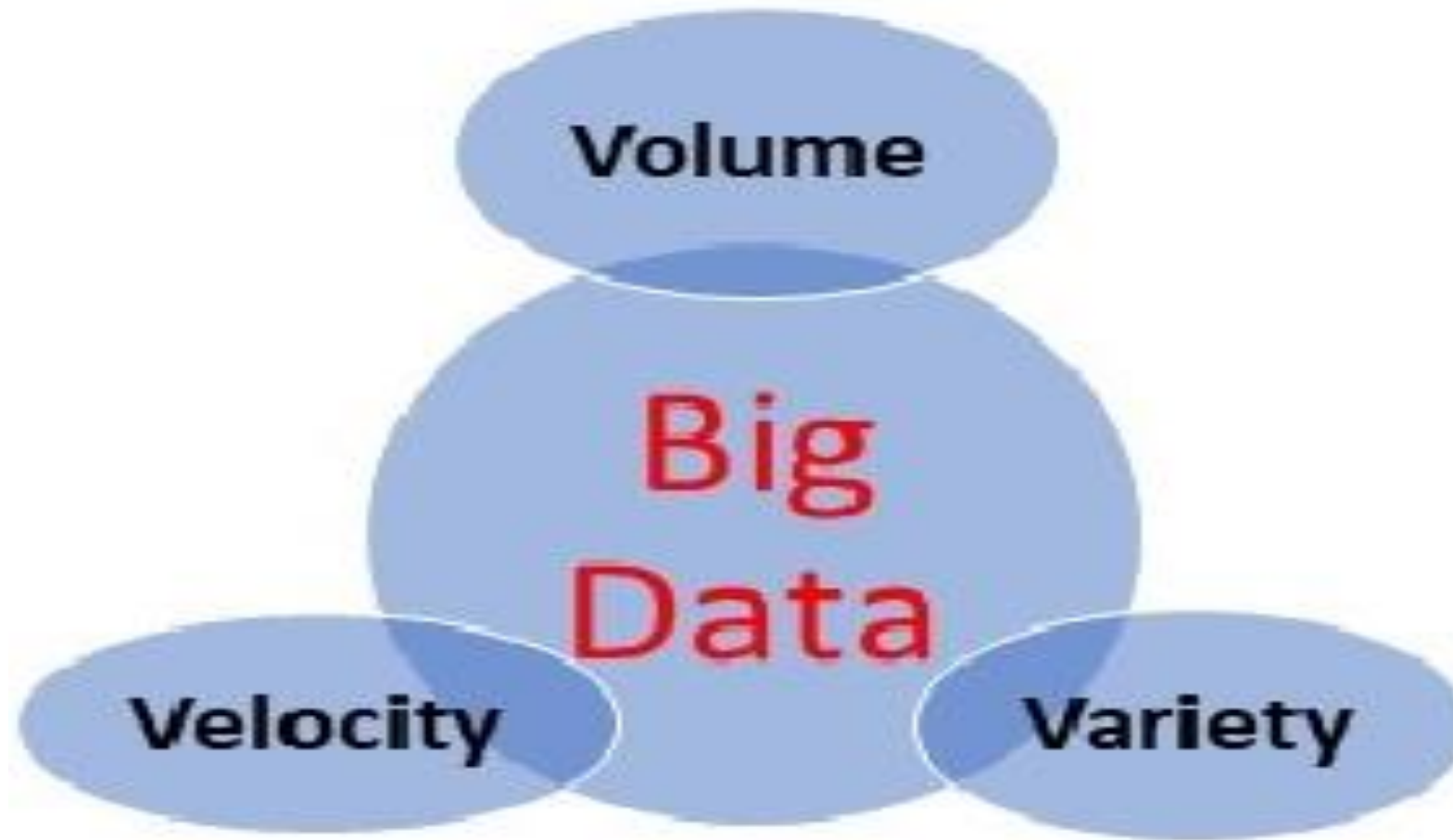
The international concepts, classifications and methods should be used by statistical agencies in each country to promote the consistency and uniformity of both national and international statistical systems.

International Cooperation

Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

Big Data

- **Big Data, as the name implies, are large sets of data that the traditional data processing and analysis software may be inadequate to handle.**
- **From the perspective of modern technological revolution, Big Data represent the digital data from our day-to-day digital activities (e.g., social media).**
- **No doubt that our day-to-day use of digital technology usually leaves footprints that offer great opportunities to official statisticians to augment or supplement the existing data sources.**
- **With the computing power now available, these digital data can be shared, cross-referenced, and used to give new statistical opportunities. Big Data could also be defined by the three 3Vs Model: volume, variety and velocity.**



- **Volume:** Big Data employs large volume of data generated by machines, networks and human interaction. In the past, the storage issue was the major challenge of large volume of data. But with decreasing storage costs, other challenges emerge, including data extraction and data analytics.
- **Velocity:** This describes the speed of data trooping from various sources. This unprecedented speed of data must be handled effectively and efficiently. The velocity is not being limited to the speed of incoming data but also the speed at which the data flows and being aggregated.
- **Variety:** Data come into the database all types of formats. Unstructured data streaming in from social media. Also, text documents, email, video, audio, log files and financial transactions.

The Use Big Data for Official Statistics

- **Big Data technologies are growing rapidly to a point where governments and organizations are prepared to adopt for their official statistics and data analytics.**
- **There are myriad opportunities to compile official statistics in new ways.**
- **According to the Big Data Project Inventory (United Nations Statistics Division, 2018) compiled by the UN Global Working Group on Big Data, 34 National Statistical Systems (NSSs) from around the world have registered 109 separate Big Data projects.**
- **Some statistical agencies are attempting to use a variety of sources, including satellite imagery, aerial imagery, mobile phone data, among others to compile a wide range of official statistical indicators.**

The Use Big Data for Official Statistics Cont'd

- The indicators include transport and tourism statistics, road safety indicators, price indices, indicators on corruption and crime, energy consumption, population density, migration, nutrition, among others.
- Improving price indices using scanner data or prices scraped from the web are by far and away the most popular projects. These approaches have been in development for many years and typically have fewer data access problems.
- The World Bank and the United Nations Global Pulse are also investing in Big Data potentials for official statistics. They have logged 91 projects on the Big Data Project Inventory and a wide variety of Big Data sources are being explored.

Challenges of Big Data for Official Statistics

As Big Data wends its way into official statistics, several challenges are also encountered. Some of the challenges, especially in Africa, are as follows:

Lack of Political Will towards Modern Data Revolution

African governments lack the political will to adequately support modern data revolution due to poor orientation on the benefits. Most of official statistical agencies are poorly managed.

The Possible Way Out

Advocacy, with an affirmative action, to the governments for renewed commitment into the modern data revolution.

Lack of Proper Understanding of Big Data

Statistical agencies in Africa fail in their Big Data initiatives due to insufficient understanding. African data professionals may know what is going on, but governments agencies may not have a clear picture.

The Possible Way Out

Big Data workshops and seminars must be held for government officials. Basic training programs must be arranged for all the stakeholders who are handling data regularly and are a part of the Big Data projects. A basic understanding of data concepts must be inculcated by all levels of the organization.

Data Growth and Storage Challenges

As the amount of data being stored in the servers and databases are increasing rapidly, it is extremely difficult to handle. Inadequacy of storage and warehousing of huge sets of data.

The Possible Way Out

To handle these large data sets, techniques such as compression and deduplication could be used. Compression is used for reducing the overall size of the data. Deduplication is the process of removing duplicate and unwanted data from a data set.

Big Data Talent Gap

Another challenge of Big Data for official statistics is a lack of homegrown data science talents across the continent. This could be because of high cost of quality education and short trainings in Data science and Statistics in Africa.

The Possible Way Out

The cost of quality education and short trainings particularly in Data science and Statistics should be subsidized by Governments and other international organizations to bridge the gap.

Difficulty of Data Synchronization and Processing

The difficulty to synchronize and possibly harmonize data sets from diverse sources and incorporated into an analytical platform. Also, the difficulty to collect, transform and process large volume of data to extract the significant information.

The Possible Way Out

This could be overcome through rigorous trainings on data handling as well as the use of modern facilities for data analytics.

Computational and Data Handling Complexities

Some of conventional techniques may not be suitable for large datasets. The available tools may not be sufficient to process these Big Data to obtain meaningful information. Handling of large dataset requires more advanced technologies.

The Possible Way Out

This could be overcome through data warehouses and data marts. Data warehouse is mainly responsible to store data that are sourced from various systems whereas data mart is data warehouse of a single system.

Legislative/ Management Challenges

Big Data for official statistics implies that more information coming to the official statistical agencies and managing them is a challenge for Africa. Also, Big Data is not adequately supported by specific laws/obligations. Many potential Big Data sources are collected by non-governmental organizations and are always freely available on the web; situations that may not be covered by existing legislation.

The Possible Way Out

Governments and private sector Big Data professionals should come together to fashion out an enabling data management and legislative frameworks for Big Data sources.

Financial Challenges and Poor Infrastructure

Potential high costs of sourcing data and poor infrastructure to handle Big Data projects. There is likely to be a cost to host projects, acquire Big Data especially those held by the private sector.

The Possible Way Out

This could be overcome through collaboration with the private sector and the provision of more funding by governments and international partners.

Conclusion

- **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.**
- **In years to come, developments such as the Internet, biometrics, among others will present other opportunities to compile new and useful official statistics.**
- **Big Data may offer new cost-effective or timely ways of compiling statistics or offer some relief to survey fatigue and burden. Big Data for official statistics has numerous challenges in Africa, ranging from lack of understanding, low capacity, funding, data handling, legislative issues, among others.**
- **These challenges could be resolved through concerted efforts from governments, private sector and the international partners.**

Recommendations

- **The benefits of using Big Data for official statistics have outweighed the challenges.**
- **African Countries through their governments and the private sector must come together to address the challenges and reap the benefits of Big Data towards enhancing official statistics in Africa.**

THANK YOU

References

- Gandhi, R.V., Kumar, R. and Vamshi, K.P. (2017). Big Data: issues and challenges. *International Journal of Software & Hardware Research in Engineering*, 5(1), 1-5.
- Johanna, G. and Katalin, S. (2021). Official Statistics and Big Data. *East European journal of society and politics*, 4 (1), 159-182.
- Piet, D. and Mark, V. (2013). Big Data and official statistics. *Meeting on the Management of Statistical Information Systems*. France and Thailand.
- Steve, M (2020). Big Data and Official Statistics. *The United Nations Conference on Trade and Development (UNCTAD)*. Switzerland.