

## Opportunities, Challenges and Building Partnerships for Official Statistics in the Era of Big Data in Nigeria.

<sup>1</sup>Osuolale Peter Popoola and <sup>2</sup> Nureni Olawale Adeboye

<sup>1</sup>Mathematics and Statistics Department, The Ibarapa Polytechnic, Eruwa. Oyo  
State Nigeria.

<sup>2</sup>Department of Mathematics and Statistics, Federal Polytechnic, Ilaro Ogun State Nigeria

Email: [osulalepeter@yahoo.com](mailto:osulalepeter@yahoo.com),

### Abstract

Official Statistics are produced, collated, and disseminated by the federal governments of Nigeria through the Nigeria Bureau of Statistics (NBS). These data are almost invariably nationally representative, because they are obtained from complete censuses or very large-scale national sample surveys, and they usually seek to present definitive information conforming to international definitions and classifications or other well-established conventions. The impersonal characters of official statistics, and their resistance to innovation, stand in sharp contrast to statistics and data-sets from other sources such as, internet, social media, academic research, market research, independent research institutes, and commercial organizations. In our modern world more and more data are generated on the web, social media and sensors in the ever growing number of electronic devices surrounding us but despite this unprecedented growth, much of the value of data is still untapped, waiting to be realized. The volumes of data and the rate at which these data are produced have led to the concept of 'Big Data'. The rise in big data should change the context in which NBS operate in Nigeria. Big Data provides opportunities to obtain timely, costless, higher precision, completeness and less burden data but in order to make optimal use of Big Data for official statistics, a number of challenges have to be addressed. This paper outlines opportunities and challenges of obtaining official statistics and presents how to build a Public Private Partnership model in obtaining official statistics in the era of big data in Nigeria.

Keywords: *Big Data, National Bureau of Statistics (NBS), Official Statistics, Mobile Network providers, National Communication Commission (NCC)*

### Introduction

When the Earth was formed about 4.5 billion years ago, the early years were pretty chaotic. Eventually, oceans formed, and simple organisms began to evolve, followed by more complex plants, animals and, finally, man. Statistics have also been evolving into more complex forms-

granted over a much shorter period of time. The earliest data collections took the form of census. Sampling was eventually discovered and gave rise to surveys and then multi-topic surveys.

The omnipresence of data in the daily lives of most people in the world gives rise and support to the view that data will change the world. With the unprecedented rate of data creation, and the increasing role data plays in most of our lives, it is easy to assume that the digital revolution could be the most important life-changing event of this era. And, as the world battles COVID-19, the value and potential development impact of data have become all the more evident. More recent years have seen a Cambrian explosion of technology: remote sensing, electronic data capture, and data from mobile phone users (Big Data). The rapid development in Information and Communications Technology (ICT) has enabled information to be generated and shared quickly nowadays. Electronic gadgets, such as cellular phones, satellites, Global Positioning Systems (GPS), and scanning devices, and fora like social media and e-commerce create volumes of data on a daily basis, and in some instances by the second. There are over 7 billion mobile phone subscriptions and 3 billion internet users worldwide. Mobile broadband subscriptions increased from 268 million in 2007 to over 2.1 billion in 2013 (ITU, 2013). The information generated by these media constitutes *data exhaust* and is defined as “the digitally trackable or storable actions, choices, and preferences that people generate as they go about their daily lives” (Global Pulse, 2012 p. 9). IBM reported that over 2.5 quintillion bytes of data are generated daily. The stock of digital data rose from 150 Exabyte in 2005 to 1200 Exabyte in 2010 (Global Pulse, 2012). This kind of fast moving, high volume data have been dubbed Big Data.

### 1.1. Big Data and Official Statistics

According to Popoola & Nuahmah (2018) define big data as data sets that are so voluminous and complex that traditional data processing application soft wares are inadequate to deal with. It is typically characterized by the so called, seven “V’s” namely; volume, velocity, variety, Value, Veracity, variability and validity. They further defined Big Data as very high volume, velocity

and variety of data that require a new high-performance processing. Big data are generated digitally and can be stored as binary data. Big Data are produced by daily interactions with digital services. Based on this, mobile phone data, Twitter, Facebook, and online queries can all be classified as big data. Big data has been promoted as a timelier and cheaper alternative to traditional sources of official data for national development. Big data has gained much attention in recent times in large part due to the fact that it is available in a timelier manner than other data and can be generated at minimal cost. The low cost to collect, store and process data can be attributed to technological advancements over the last two decades. New sources of data become available every day through mobile phones, web searches, purchases etc. Online information is sourced from social media such as Twitter and blogs, news media, job postings, etc. Online information creates a sensor of human intent, sentiments, perceptions, and wants while physical sensors detect changes in human activity. Such changes can be indicated by traffic patterns, light emissions, and satellite imagery of changing landscapes or urban development. Crowd sourcing uses the crowd as a free source of information. Such information is provided through mobile phone-based surveys or hotlines and is used for verification and feedback purposes. These unstructured and complex data can prove to be invaluable input in governance and policy making and there have been developments in methodologies such as advanced computational tools to enable the exploitation of big data for these and other purposes. The question to be addressed is Should National Bureau of Statistics (NBS) the body that produces and coordinates official statistics for Government in Nigeria actively engage in capitalizing opportunities that big data now affords or should they become a third party that certifies data from private sources? There is a possible opportunities for NBS in collecting non-traditional data and to use them to complement traditional sources of official statistics in Nigeria. For instance, conducting a new survey could cost over ₦100 million in the Nigeria. Using big data can produce similar results at one-tenth of this cost.. In this era of big data revolution, big data could fill the gaps in traditional data collection since traditional national household and business surveys were losing popularity, big data could provide timely and relevant statistics (Landefeld, S. 2014). The fact that the private sector had a pivotal role in providing big data yet enough was not being done to fully engage and incorporate them in the process of using big data for official statistics in Nigeria.

## **1.2. National Bureau of Statistics and data collection in Nigeria**

Nigeria operates Federal System of government with 36 States and Federal Capital Territory and 774 Local Government Areas (LGAs). At the federal level, each Ministry, Department and Agency has Director of Statistics. Each state has Director of Statistics and Head of statistics Unit at Local Government Areas. All these including Statistical Institutes constitute the Nigeria National Statistical System (NSS). The National Bureau of Statistics (NBS) came into being with the merger of the Federal Office of Statistics (FOS) and the National Data Bank (NDB). The creation was part of the implementation of the Statistical Master Plan (SMP), a program document of the Federal Government of Nigeria (FGN). The merger was to give the agency a national outlook as the apex statistical agency for all the three tiers of Government. NBS is expected to coordinate Statistical Operations of the National Statistical System in the production of Official Statistics in all the Federal Ministries, Departments and Agencies (MDAs), State Statistical Agencies (SSAs) and Local Government Councils (LGCs). (Extracted from the website of NBS).

## **2.0. Big Data for Official Statistics: The Opportunities for NBS**

The benefits of big data are still emerging but there are already clear indications that big data could enhance timeliness in the production and dissemination of information while being more cost-effective than traditional means.

There is a rapid increase in the volume of data that is made available on a daily basis via the web, mobile devices, and the Information Technology structure. The number of device users and digitization of transactions have increased the speed at which such data are created. Data are not represented in the traditional format but are generated instead in unstructured and semi structured forms. Big data has been promoted as a timelier and cheaper alternative to traditional sources of official data. Let us first look at the opportunities for NBS offered by various big data sources in Nigeria.

### **2.1. Internet(Machine Generated Data)**

The internet is a very important part of our day to day life in the 21st century. It helps us to access information, have the world at our reach, and helps to connect the whole world in one place. It is important for business enterprises, students, and individuals. This includes data from

sensors, home automation, weather sensors, and traffic sensors, mobile phone location, car sensors and satellite images. For instance, traffic light sensors are installed in the major road in almost all the major cities in Nigeria. Data on traffic and transport statistics can be obtained through this medium without conducting any statistical survey. Data collection by questionnaire on road usage may not be necessary anymore if detailed traffic loop data from sensors in roads, become available.

## **2.2. Social Networks**

Nigeria has Africa's largest mobile market, with about 173 million subscribers out of about 200 million populations. The initial rapid growth in the number of subscribers had led to problems with network congestion and quality of service, prompting the regulator to impose fines and sanctions on network operators. These operators have responded by investing billions of dollars in base stations and fiber transmission infrastructure to support the increasing demand for data. The migration from the CDMA platform to GSM technology is almost complete, and though GSM still dominates the market there is a growing shift to services based on LTE. The use of mobile phones in Nigeria is ubiquitous. People often carry phones with them and use their phones throughout the day. Instrumental for the infrastructure enabling the coverage for mobile phones, are mobile phone masts/towers, called 'sites' in the industry. Those sites are located at strategic points in Nigeria, covering as wide an area as possible. Almost all the mobile phone companies such as: MTN Nigeria, Glo Mobile (Globacom), Bharti Airtel (formerly Zain, Celtel), 9Mobile (Etisalat Nigeria, EMTS, Mubadala), Vodacom etc. record data that are very closely associated with behavior of people; behavior that is of interest to statistical agencies. Obvious examples are behavior regarding tourism, mobility, commuting and transport. The destinations and residences of people during day-time are topics of various surveys. Using data from mobile phone companies this should be able to provide additional and more detailed insight on the whereabouts and the activity of mobile phone users. Location data for mobile phones could be used for almost instantaneous daytime population and tourism statistics (De Jonge et al., 2012).

## **2.3. Traditional Business systems (process-mediated data)**

Data produced by public agencies, businesses commercial transactions, banking/stock records, E-commerce; credit cards etc. could be harnessed by the NBS. Mobile positioning data provide very accurate and precise geographical information. Therefore, they can be useful in providing mobility statistics and in conducting spatial behavioral analyses. Mobile positioning data collection is relatively easy to set up and is not demanding of the respondents. The drawback to using mobile positioning data lies in the limits to the requests that can be made via mobile network operator's network infrastructure and the possible small sample size that could result from the need to seek permission from subscribers (Tiru, 2014). Mobile data have been used successfully in tracking persons in Haiti after the 2010 earthquake using the movement of two million SIM cards. Health information collected through similar means by health workers or provided by the individuals themselves can be useful in predicting and stopping disease outbreaks. The daily movements of persons can be tracked using the data provided by mobile devices. It represents a "more reliable, accurate, timely, and less expensive alternative to traditional methods" to establish "travel patterns between home and work-time locations" (Tiru, 2014 p. 14). Also, mobile data can be used in the areas of spatial segregation, emergency and public safety, social networking analysis and mobility, and climate change. Other sources of big data are equally important. Satellite imagery has great potential, particularly for the agricultural sector. Around 16 million text messages are produced on a daily basis in the Nigeria. These messages are available to anyone with internet access. Social media has the potential of being a data source as people voluntarily share information; discuss topics of interest, and contact family and friends. Social media messages could be used for several types of indicators, such as an early indicator of consumer confidence. Inflation figures could be derived from price information on the web, and so on.

### **3.0. Challenges of Incorporating Big Data for Official Statistics in Nigeria**

As much as big data could be used for official statistics in Nigeria, it comes with many challenges as well. These are mainly related to methodology, privacy and legal concerns, access, technology, and capacity.

(a) *Legislative*, i.e. with respect to the access and use of data. The right to access admin data, established in principle by the law, is not adequately supported by specific obligations for big

data. Many potential Big Data sources are collected by non-governmental organizations or are 'freely' available on the web; situations that may not be covered by existing legislation.

(b) *Privacy*, i.e. managing public trust and acceptance of data re-use and its link to other sources.

Privacy is generally defined as the right of individuals to control or influence what information related to them may be disclosed. The problem with Big data is that the users of services and devices generating the data are most likely unaware that they are doing so, and/or what it can be used for. The data would become even bigger if they are pooled, as would the privacy concerns.

(c) *Financial*, i.e. potential costs of sourcing data vs. benefits. There is likely to be a cost to acquire Big data, especially Big data held by the private sector and especially if legislation is silent on the financial modalities surrounding acquisition of external data.

(d) *Management*, e.g. policies and directives about the management and protection of the data.

Big data for official statistics means more information coming to NBS that is subject to policies and directives on the management and protection of the information to which NBS must adhere.

Long-term stability may be a problem when using Big Data. Typically, statistics for policy making and evaluation are required for extended periods of time, often covering many years.

Many big data sources have only recently been 'established'.

(e) *Methodological*, i.e. data quality and suitability of statistical methods. When more and more data are being analyzed traditional statistical methods, developed for the very thorough analysis of small samples, run into trouble; in the most simple case there just not fast enough. Since text is an essential part of many Big Data sources, the need to extract information from text increases. Also, the subpopulations covered by Big Data sources studied are not the target populations for official statistics. Therefore such data are likely to be selective, not representative of a relevant target population. Assessing representatively of big data may prove problematic, as often there are no characteristics readily available to conduct such comparison.

(f) *Technological*, i.e. issues related to information technology. Dedicated and specialized computing infrastructures are required to cope with Big Data to enable processing and speed up analysis of large amounts of data. Certainly for the exploratory phase, during which the content and structure of big data sets has to be understood, fast technology certainly speeds up this process and more quickly enable the revelation of their use for statistics.

#### **4.0. Building a Public, Private Partnership model in obtaining Official Statistics in era of Big Data in Nigeria.**

The question to ask is how could National Bureau of Statistics (NBS) build a public private partnership model to obtain official statistics taking advantages offer by various big data sources such as internet, social network providers, IT companies, public agencies etc. in Nigeria? There is need to develop agenda that will require partnership of NBS in using traditional data sources with the nontraditional data producers in Nigeria– ranging from civil society, mobile network providers, IT companies, academia and to research institutions in Nigeria. The below model may be employed:

##### **4.1. NBS forming partnership with other Agencies of Government and other parties**

National Bureau of Statistics could form a partnership with third parties such network service providers or IT/consulting companies with experience in the statistical analysis of large data sources. There are institutions that are already collecting the type of information that will be relevant to national development agenda. For example, the initiatives of the World Resources Institute have made available information on food security, ecosystems and human well-being. There is a need to develop plans on how such data can be harnessed from the various private and public sources. First of all, the potential providers of Big Data such as network providers (MTN Nigeria, Glo Mobile (Globacom), Bharti Airtel (formerly Zain, Celtel), 9Mobile (Etisalat Nigeria, EMTS, Mubadala), M-Tel (Nitel), Visafone, Starcomms (Capcom), Multi-Links, Reliance, InterC Network (Intercellular), etc. are essential partners if they do not grant access to their data, the story is over before it starts. In Nigeria, the Nigerian Communications Commission (NCC) is the independent National Regulatory Authority for the telecommunications industry. The Commission is responsible for creating an enabling environment for competition among operators in the industry as well as ensuring the provision of qualitative and efficient telecommunications services throughout the country. Consistent with Section 89 Subsection 3(d) of the Nigerian Communications Act 2003 (NCA 2003), the Commission is mandated to monitor and report on the state of the Nigerian telecommunications industry, provide statistical analyses and identify industry trends with regard to services, tariffs, operators, technology, subscribers, issues of competition and dominance, etc. with a view to



identifying areas where regulatory intervention would be needed; In view of this, the Commission regularly conducts studies and surveys and produces reports on the telecommunications industry, and telecommunications operators are obligated, under the terms of the licenses, to provide the NCC with such data on a regular basis for analytical review and publishing. National Bureau of Statistics (NBS) can form partnership with NCC (both are Government Agencies) and other IT company in Nigeria to obtain official statistics on various behavior of the population without using traditional method of drafting of questionnaire to obtain qualitative and quantitative data. Another way NBS could partnership with network service providers is that of a trusted third party. In a competitive market, competitors will be reluctant to share sensitive data among each other. But they might be willing to share it with an NBS who compiles statistical information that is beneficial to all. Partnership between NBS and academia may grow as well. Universities have historically been natural partners with NBS. It stands to reason that such Partnership will extend to the field of Big Data, for instance, in solving methodological problems, developing technical solutions and training future data scientists. Such partnership is also being supported by public funders who are facilitating research and innovation partnerships through targeted grants. By working in partnership, researchers in universities and NBS could better leverage such opportunities. Furthermore, there are many commercial partners with which NBS could partner. Google and Facebook are two examples for which Big Data forms the core of their business model. Their knowledge and the data to which they have access may be very relevant to NBS. IT companies also possess relevant knowledge on Big Data processing and storage, security, cloud processing, etc. Apart from the provision of paid services, partnership may be of interest to them with a view to obtaining statistical expertise and for benchmarking or validating their information products.

#### **4.2. Develop a new exploration and analysis Techniques**

NBS could develop a new exploration and analysis methods, specific for the study of huge volumes of data, in the context of official statistics. Various potential producers of big data have their own concerns and, like NBS, they are subject to privacy rules. This may complicate partnership even if they have a positive outlook and approach. But since big data sources are not designed for statistical use, such partnership is also essential in order to obtain good knowledge

of the provenance of such sources. Additionally, for statistical production, it may be more efficient to have data processed at the site of collection and storage. In such cases, the assumption that data can be provided for free may no longer hold. On the other hand, statisticians also have much to offer such as providing analytic insights that may help data owners understand their data better. Doing complex statistical analyses are core business for NBS, but not for, say, mobile network providers. In these and other ways, the relationship with data providers could potentially become true partnerships if NBS could develop a new exploration and analysis methods to study huge volumes of data being generated from these third parties.

#### **4.3. Experiment with Big Data**

NBS can Experiment with Big Data sources by setting up a number of pilot projects in selected in statistical areas. These pilots will provide guidelines for the effective use of Big data for purposes of official statistics. Important research areas may include: combining Big Data with traditional data sources (survey, administrative); replacing traditional data sources, i.e. decreasing administrative burden; opportunities for new output; opportunities for faster or real-time statistics production.

#### **4.4 Development of a Big Data training program**

Training of NBS staff is essentials if the NBS is serious about obtaining official statistics in the era of big data revolution in Nigeria. Big data is a highly multidisciplinary field requiring subject matter knowledge, strong math *skills as* well as strong programming skills. To ensure a speedy progress, research subprojects should be performed by small, highly skilled and dedicated teams covering such expertise. Moreover, because of the multidisciplinary character of the research program, guidance could be provided by a steering committee composed of experts in various statistical fields. Most staffs of the NBS are used to obtain official statistics data through the traditional method of sample survey. Statistical agencies engaged in big data projects know the importance of collaborating with abundant human and technical resources to proficiently exploit big data. The UN Global Working Group concluded that a multidisciplinary project team from different professional backgrounds is necessary to adapt the analytics to the data—rather than relying on traditional collection means such as carefully designed questionnaires. Statistical agencies, central banks, public agencies, and international organizations will not only have to

train and develop existing staff to deal with big data but also must compete with the private sector to recruit or contract new staff. These new staff members must be familiar with big data—for example, data scientists, IT architecture specialists, and data visualization specialists who work alongside subject matter professionals. On average, a big data practice in the initial stages starts with a core team of three to four staff members with a mix of technical and business skills..

#### **4.5 Development of a communication strategy**

NBS needs to develop a new communication strategy with other government agencies, private organizations, and various third parties that are producing big data so as to make it easier to obtain official statistics from their day to day administrative work.

#### **5.0. Conclusion**

The prevalence of the internet, mobile devices and other technological gadgets, a new type of data has emerged that has an advantage over traditional data sources in its timeliness, cost efficiency and other consideration. In this big data era, data are no longer centralized, highly structured and easily manageable, but are highly distributed, loosely structured (if structured at all), and increasingly large in volume. However, to successfully incorporate big data into official statistics in Nigeria, it is essential to drive legislative changes that would provide a legal framework for accessing, processing and utilizing big data in official capacities by the federal, state and local government of Nigeria; National Bureau of Statistics needs to work in partnership with other government data producing agencies, private and private institutions in Nigeria. Privacy must be maintained when acquiring, storing, and using big data. There is also the need for a business model that outlines roles and responsibilities, beneficiaries, and funding arrangements for a successful big data initiative between NBS and other partners. This calls for increased partnership of NBS with private and academic partners who have access to specific Big Data sources and knowledge. The relationship between the various stakeholders will involve each partner building on and contributing different strengths and will likely result in flexible networks. Big data is a potential source of information that cannot be ignored by official statisticians in Nigeria and that official statisticians must organize and take urgent action to exploit the possibilities of harness the challenges effectively. Conclusively, incorporating big data sources into production of official statistics in Nigeria can be better positioned to obtain

official statistical data on the economy, financial, social and the environment in terms of improved timeliness and cost efficiency, and a lessened resource burden.

**References:**

**Daas, PJH, Puts, MJH (2014a):** Big Data as a source of statistical information. *The Survey Statistician* 69: 22–31.

**Daas PJH and Puts MJH (2014b):** Social media sentiment and consumer confidence. Paper for the workshop on using Big Data for forecasting and statistics, Frankfurt, Germany, 7–8 April.

**Daas PJH, Puts MJ, Buelens B, et al. (2013):** Big Data and official statistics. Paper for the 2013 NTTS conference, Brussels, Belgium, 5–7.

**De Jonge E, Van Pelt M and Roos M (2012):** Time patterns, geospatial clustering and mobility statistics based on mobile phone network data. Discussion paper 2012, 14, Statistics Netherlands.

**Global Pulse (2012):** Big Data for Development: Challenges & Opportunities, <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>.

**ITU (2013):** Big Data: Big Today, Normal Tomorrow, [http://www.itu.int/dms\\_pub/itu-t/oth/23/01/T23010000220001PDFE.pdf](http://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000220001PDFE.pdf).

**Landefeld, S. (2014):** Uses of Big Data for Official Statistic: Privacy, Incentives, Statistical Challenges and Other Issues. Discussion Paper, International Conference on Big Data for Official Statistics, Beijing China, 28 – 30 October. <http://unstats.un.org>.

**Osuolale Peter Popoola & Nicholas Nsowah Nuamah (2018):** New Trends in Modeling Climate Change in the Era of Big Data, *Anale. Seria Informatică*. Vol. XVI fasc. 2 – 2018

**Struijs P and Daas PJH (2013):** Big Data, big impact? *Paper presented at the seminar on statistical data collection*, Geneva, Switzerland, 25–27 September 2013.

**Tiru, M. (2014):** Overview of the Sources and Challenges of Mobile Positioning Data for Statistics, International Conference on Big Data for Official Statistics, Beijing. <http://unstats.un.org/unsd/trade/events/2014/beijing>.