



Nations Unies  
Commission économique pour l'Afrique

## **Report on Fostering African Private Sector in the Big Data Era**

**Technology, Climate Change and Natural Resource Management Division**

May 2020

## Contents

1. Introduction
  2. Big Data and the Data Revolution
    - 2.1 What is Big Data?
    - 2.2 The 5Vs of Big Data
    - 2.3 Big data and open data
  3. The Big Data Ecosystem
    - 3.1 The Big Data Value Chain
    - 3.2 The stakeholders in the Big Data Ecosystem
    - 3.3 The Role of Stakeholders
  4. Big data and the Private Sector
    - 4.1 Participation of the African private sector in the Big Data era
      - 4.1.2 The growth of tech hubs
      - 4.1.2 The use of big data in development
      - 4.1.3 Business opportunities in Big Data use in government
    - 4.2 Use of big data for competitive advantage of companies
  5. Challenges
  6. Conclusion and Recommendations
    - 6.1 Conclusions
    - 6.2 Recommendations
- Annexes
- Annex I Types of Big Data
- Annex II Selected companies in Africa engaged in open data based business

## 1. Introduction

In this globally networked world, people, objects and connections are producing data at unprecedented rate, both actively and passively. This phenomenon not only creates large volumes of data but also distinctive streams of data that have been termed 'big data'. Big data refers to emerging technological capabilities in solving complex tasks which is considered a new frontier for innovation, competition and productivity.

According to Ericsson, more than 50 billion devices will be connected by 2020 irrespective of what and where they are, enabling an Internet of Things (IoT) and Places. These devices are there to measure, sense, generate and communicate data of some size and structure. ITU estimates that there are almost 7 billion mobile-cellular subscriptions worldwide, and each of them is both a data creator and consumer. Currently over half of the world's population use the Internet and mobile-broadband subscriptions. Each of these users contribute to the data revolution, with SMS, calls, photos, videos and messages posted on social media sites, e-mails, searches, clicks on links and ads, online shopping and mobile payments, or location traces left from GPS-enabled smart phones and WiFi network logins.

Researches show that there are several benefits associated with Big Data initiatives particularly in the context of private sector ranging from better, fact-based and fast decision making to improved customer experience, increased sales, new product innovations, reduced risk, more efficient operation and higher quality products and services. As with the case of business intelligence (BI) initiatives, Big Data systems have been used for two purposes, namely, human decision support and decision automation. In this regard, the private sector holds data with a potential impact many times that of government data. Retailers, financial organisations, telecommunications companies, social network providers and other online platforms (such as search engines) hold much bigger, diverse and deeper datasets. Indeed, many of these are referred to as 'Big Data', requiring different data analysis tools and skills, whose potential for public good is seemingly unlimited. The issue therefore has been with the appropriate use of the data which led to the evolvement of the 'responsible data' movement that discusses issues of guidelines and frameworks to ensure ethical principles for data sharing.

In this regard, the use of data collected by the private sector for the public good is being practiced. For instance, Orange's 'D4D challenge' using real CDR data (mobile phone Call Detail Records) from Cote d'Ivoire and Senegal are some of the examples in this context. The D4D Challenge for Senegal, for example, demonstrated how CDR could be used in a range of sectors such as agriculture (4 use cases), energy (1), health (12), national statistics (9), transportation and urbanization (22) and eight other use cases. Such a practice is called data

philanthropy, although it also encompasses the practical and technical support needed for data analysis. The debate of using private data for public good has been on the public with issues such as the use or not of CDRs for tracking the origin and spread of Ebola, etc. To this end, there is a need for further examination of the issue of privacy, intellectual property and legal and regulatory frameworks with access and use of data is a key issue that his study considers as part of the challenges in big data developments.

Africa faces among others challenges of the lack of critical data for regional and national development policymaking. Large data gaps remain in several development areas couple with poor data quality, lack of timely data and lack of disaggregated data on important dimensions are among the major challenges. It is estimated that as many as 350 million people worldwide are not covered by household surveys. There could be as many as a quarter more people living on less than USD1.25 a day than the current estimates suggest, because they have been missed out of official surveys<sup>1</sup>.

In this context, at UNECA, the New Technology and Innovation Section, as part of its policy research activities, aims to explore the trends in growth of the ‘Big Data’ phenomenon in Africa with particular focus on the role of the private sector in the generation, use and dissemination of data for the socio-economic transformation of the continent. To this end, this report has been prepared under the UNECA on the topic ‘fostering the private sector in the big data era’ and will be further discussed during the Expert Group Meeting on the topic planned to be held from 3 to 6 December 2019 in Yaoundé, Cameroon. The study report will be further enriched from the discussions, inputs and contributions of the experts and final report will be prepared that will incorporate key policy recommendations for considerations by African policy- and decision makers.

## **2. Big Data and the Data Revolution**

The report of the UN Secretary General’s Independent Expert Advisory Group (IEAG)<sup>2</sup> “defines the data revolution for sustainable development as the integration of data coming from new technologies with traditional data in order to produce relevant high-quality information with more details and at higher frequencies to foster and monitor sustainable development. This revolution also entails the increase in accessibility to data through much more openness and transparency, and ultimately more empowered people for better policies, better decisions and greater participation and accountability, leading to better outcomes for the people and the planet”.

---

<sup>1</sup> Carr-Hill, R. (2013) ‘Missing millions and measuring development progress’, World Development 46: 30-44.

<sup>2</sup> United Nations (2014), p.6. A world that Counts: Mobilizing a Data Revolution for Sustainable Development by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. New York.

## 2.1 What is Big Data?

With the advancement of technology, the world is experiencing a data revolution, or ‘data deluge’. Today, a massive amount of data is being generated and flowing from various sources, through different mediums, every second in today’s digital world. On one hand the speed and frequency with which data is generated and transmitted, and on the other hand the rise in the number and variety of sources from which it originates, that makes the data revolution. Big data represents an umbrella term for the explosion in the quantity and diversity of high frequency digital data. Big data is therefore a massive volume of both structured and unstructured data that is so large that it is difficult to process in the traditional data processing tools and methods cannot adequately process them. The term big data or ‘big data analytics’ is defined by Gartner<sup>3</sup> as:

*“Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.”*

## 2.2 The Five Vs of Big Data

Big data is generally characterised as data with increasing volume, velocity and variety, known also as the 3Vs (Laney, 2001)<sup>4</sup> as described below.

- **Volume (amount of data)** refers to the exponentially increasing magnitude of data. The size of Big Data is expressed in terabytes and petabytes that are driven by emerging new data collection tools, such as social media, mobile applications or sensors. The amount of this collected information requires improved storage systems and data management technologies.
- **Velocity (speed of data)** refers to the speed at which Big Data are created, spread, stored, analysed and visualized. The data are virtually produced in real-time, e.g. internet logs or location tracking by mobile phones. This higher frequency of data creates pressure to increase the current speed of data processing. It also refers to the dealing of high frequency of incoming real-time data (e.g. sensors, pervasive environments, electronic training, Internet of Things (IoT)).
- **Variety (range of data types/sources):** refers to Big Data’s heterogeneous structure and the complexity of its formats. Big Data can appear in a structured pattern, such as databases, as well as in semistructured or unstructured documents, images, video, emails, coordinates, etc. Analytical tools are being continuously improved and they are now able to deal with very heterogeneously structured data. Such data do, however, consume more time and storage.

---

<sup>3</sup> Big Data definition, Gartner, Inc. Available: <http://www.gartner.com/it-glossary/big-data/>

<sup>4</sup> Laney, Douglas. 3-D Data Management: Controlling Data Volume, Velocity and Variety. META Group Reserach Note, 2001.



## 2.2 Types of Big data

Big data represents a complicated structure of various datasets involving structured and unstructured forms. Two different classifications of Big Data exist in official statistics, the UNECE’s classifications of Big Data by type<sup>6</sup>, and classifications by size<sup>7</sup>.

### 2.1.1 Classification by type

Big data are represented in various forms and derived from various sources. An internationally agreed classification of the type Big Data was developed by UNECE’s Task Team on Big Data in June 2013 as shown in the table below (Table 1). The first section of the classification consists of data derived from social networks (such as facebook, twitter, blogs, youtube, etc.) that provide user-generated data reflecting various human experiences mostly unstructured and unregulated and are characterized by high frequency. In the second classification are data derived from traditional business systems including administrative data types produced by public entities on one hand data produced by businesses on the other hand. These data are highly structured and usually include reference tables, relationships and metadata to show the context. However, the challenge with this type of data could be access to proprietary and confidential data, and the nature of the process utilized in generating this data including sample, methodology, etc. but is the most important source of Big Data. The third category of is classified as the Internet of Things (IoT) represented by machine generated data through various sensor data (e.g. on weather, pollution, traffic, etc.), mobile phone location records and satellites imagery, etc. This type of data are derived from the exponentially growing devices, sensors and machines that are measuring, recording events and conditions in the physical world.

**Table 1: Classification of Big Data by Type**

Data Source	Data Type
Social Networks	1100. Social Networks: Facebook, Twitter, Linkedin, etc. 1200. Blogs and comments 1300. Personal documents 1400. Pictures: Instagram, Flickr, Picasa etc. 1500. Videos: Youtube etc. 1600. Internet searches on search engines (e.g. Google) 1700. Mobile data content: text messages 1800. User-generated maps 1900. E-Mail
Traditional Business Systems	21. Data produced by Public Agencies Administrative data

<sup>6</sup> UNECE. Classification of Types of Big Data. 2017.

<https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data>

<sup>7</sup> Dornik, Jurgen A., and David F. Hendry. "Statistical model selection with "Big Data"." *Cogent Economics & Finance* 3, no. 1 (2015).

	<ul style="list-style-type: none"> <li>22. Data produced by businesses <ul style="list-style-type: none"> <li>2210. Commercial transactions</li> <li>2220. Banking/stock records</li> <li>2230. E-commerce</li> <li>2240. Credit cards</li> </ul> </li> </ul>
Internet of Things (IoT) (machine-generated data)	<ul style="list-style-type: none"> <li>31. Data from sensors <ul style="list-style-type: none"> <li>311. Fixed sensors <ul style="list-style-type: none"> <li>3111. Home automation</li> <li>3112. Weather/pollution sensors</li> <li>3113. Traffic sensors/webcam</li> <li>3114. Scientific sensors</li> <li>3115. Security/surveillance videos/images</li> </ul> </li> <li>312. Mobile sensors (tracking) <ul style="list-style-type: none"> <li>3121. Mobile phone location</li> <li>3122. Cars</li> <li>3123. Satellite images</li> </ul> </li> </ul> </li> <li>32. Data from computer systems <ul style="list-style-type: none"> <li>3210. Logs</li> <li>3220. Web logs</li> </ul> </li> </ul>

Source: UNECE (2017)

### 2.1.2 Classification by Size

The classification of Big Data can also be based on their numerical features, such as the size of data sets<sup>8</sup>. Accordingly, three types of big data based on size can be distinguished, namely tall, fat and huge. Based on the dimensions of a data set, the number of variable  $D$ , and the number of observations  $N$ , a ‘tall’ data set can be represented by  $N \gg D$ , i.e., which doesn’t entail all too many variables, but a high number of observations, namely, a cross-section of observations of many individuals at a single point in time, for example, financial transactions or search queries. The ‘Fat’ data set represents data sets that contain high number of variables but not so many observations,  $D \gg N$ . This data type is referred to as high-dimensional data, i.e. the dimension may in fact be larger than the sample size. It is therefore represented by a time series of observations on many variables. Actually, such high-dimensional data are growing exponentially, like several statistical and other data science methods. Often used techniques for such data are machine learning methods, dimensional reduction techniques or Bayesian methods. Finally, Big Data sets are ‘Huge’ when they represent many variables and many observations, namely a very large number  $D$  and  $N$ . Such type of data is also known in econometrics as multidimensional panel data.

---

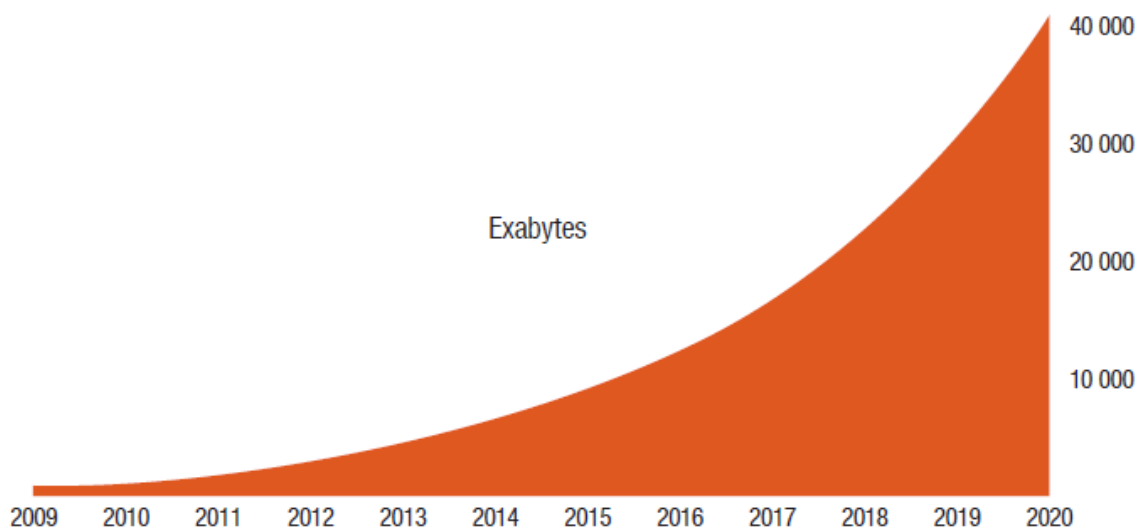
<sup>8</sup> Ibid 5



Just to show the growing size of big data (as shown in the figure below), the last decade saw an exponential growth.

It is forecasted the data will grow exponentially from around 3 zettabytes in 2013 to about 40 zettabytes by 2020 as shown in the figure below<sup>9</sup>. Big data enable values to be created in new ways and insights to be made at higher level affecting organizations, markets and government-citizens interactions. The collection and analysis of big data can be used proactively for various purposes ranging from administrative to commercial goals or passively through what is called the digital exhausts of the World Wide Web (web pages and social media), sensor-based devices and data logs generated by computing devices.

Figure 2: Growth of big data 2009 to 2020



Source: IDC (2012)

<sup>9</sup> It is estimated by the International Data Corporation (IDC) of all the digital data created, replicated and consumed in a single year. Examples of data included in the estimate consist 'images and videos on mobile phones uploaded to YouTube, digital movies populating the pixels of our high-definition TVs, banking data swiped in an ATM, security footage at airports and major events such as the Olympic Games, subatomic collisions recorded by the Large Hadron Collider at CERN, transponders recording highway tolls, voice calls zipping through digital phone lines, and texting as a widespread means of communications' (IDC, 2012). An exabyte is 1,000,000,000,000,000 bytes, and a zettabyte is 1,000 exabytes.

## 2.3 Big Data and Open Data

In the context of fostering the private sector in the big data era, it is worth elaborating the interface between big data and the emerging phenomena of ‘open data’. They are both closely related but are not the same. According to the International Open Data Center Charter (ODC) open data is ‘*publicly available data that can be universally and readily accessed, used and redistributed free of charge. It is structured for usability and computability*’<sup>10</sup>.

Open data definitions include two basic features: the data must be publicly available for anyone to use, and it must be licensed in a way that allows for its reuse.

Furthermore, some although not nearly enough useful data collected by the private sector has been opened up under a number of different arrangements, such as through what is called data philanthropy which will be further addressed later.

A subset of open data is Open Government Data (OGD) which is open data generated and released by local or regional Government ministries, departments and agencies (MDAs). Mostly, the National Statistical Office (NSO), with its mandate to collect, approve the quality of, and release official statistics, is also the key player in releasing OGD. However, in many countries there is a distinct government organ for this purpose. Moreover, many different OGD technical platforms, URLs and interfaces are used to host or access the open data portals. Besides OGD, other open datasets are created or curated by different stakeholders in the data ecosystem, such as local and international NGOs, local governments, academic institutions and private organisations.

## 3. The Big Data Ecosystem

Exploiting the biological metaphor of ecosystem, a business ecosystem can be defined as ‘economic community supported by a foundation of interacting organisation and individuals’<sup>11</sup>. This expands the need for businesses to collaborate among themselves to survive within a business ecosystem. In a business ecosystem, a smart company manages information and its flows. Therefore, in terms of data, the ecosystem metaphor is useful to describe the data environment supported by a community of interacting organisations and individuals. Big Data Ecosystems can be formed in different ways around organisations, community, technology platforms, or within or across sectors. For instance, sectors that have established or emerging

---

<sup>10</sup> <https://opendatacharter.net/>

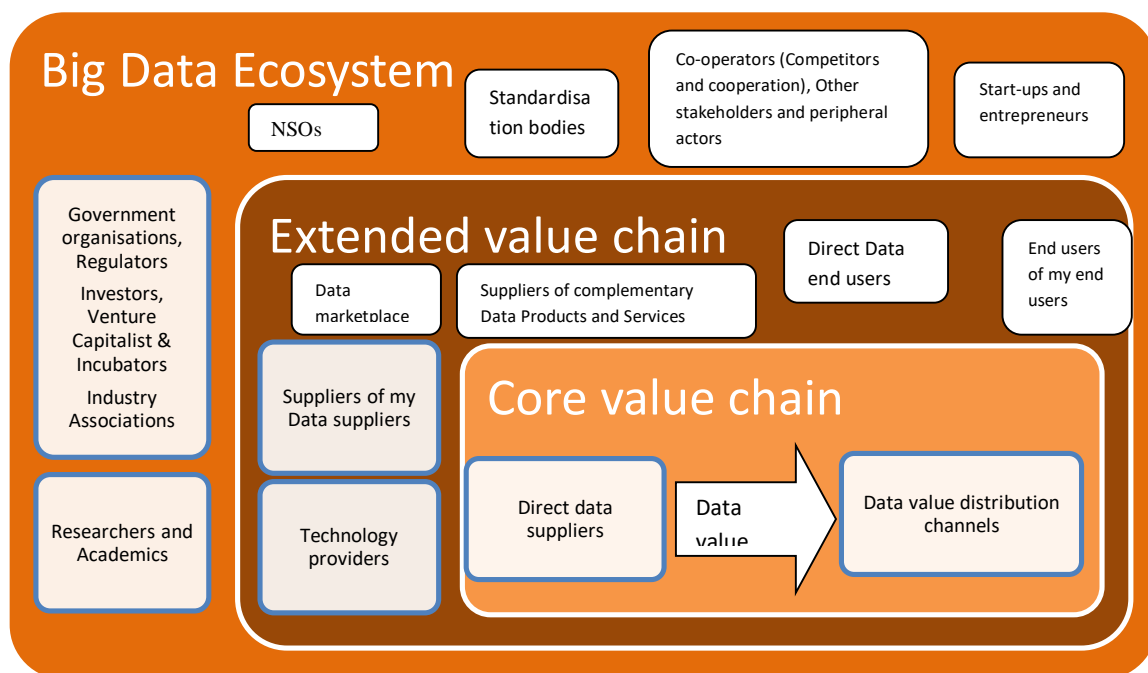
<sup>11</sup> Kim, H., Lee, J.-N., & Han, J. (2010). The role of IT in business ecosystems. *Communications of the ACM*, 53, 151.

data ecosystems include Healthcare, Finance<sup>12</sup>, Logistics, Media, Manufacturing and Pharmaceutical<sup>13</sup>. Furthermore, the Big Data Ecosystem can also be supported by other actors including data management platforms, data infrastructure and data services.

As opposed to other sectors, the stakeholders in the Big Data arena are widely dispersed not yet well connected and some processes need to be in place to bring them together. Exploiting the potential of big data needs collaboration of various actors including data scientists and practitioners to put insights generated from big data in to use.

In this regard, a well-functioning and working data ecosystem should bring together key stakeholders with clear benefits for all. The key actors in a big data ecosystem are described in Figure 3 below:

**Figure 3: The Micro, Meso and Macro Levels of a Big Data Ecosystem**



Source: Adapted from Moore (1996)<sup>14</sup>

- **Data Suppliers:** Individuals or organisations (large, small and medium-sized enterprises (SMEs) that create, collect, aggregate, and transform data from both public and private sources

<sup>12</sup> O’Riain, S., Curry, E., & Harth, A. (2012). XBRL and open data for global financial ecosystems: A linked data approach. *International Journal of Accounting Information Systems*, 13, 141–162.

<sup>13</sup> Curry, E., Freitas, A., & O’Riain, S. (2010). The role of community-driven data curation for enterprises. In D. Wood (Ed.), *Linking enterprise data* (pp. 25–47). Boston, MA: Springer US.

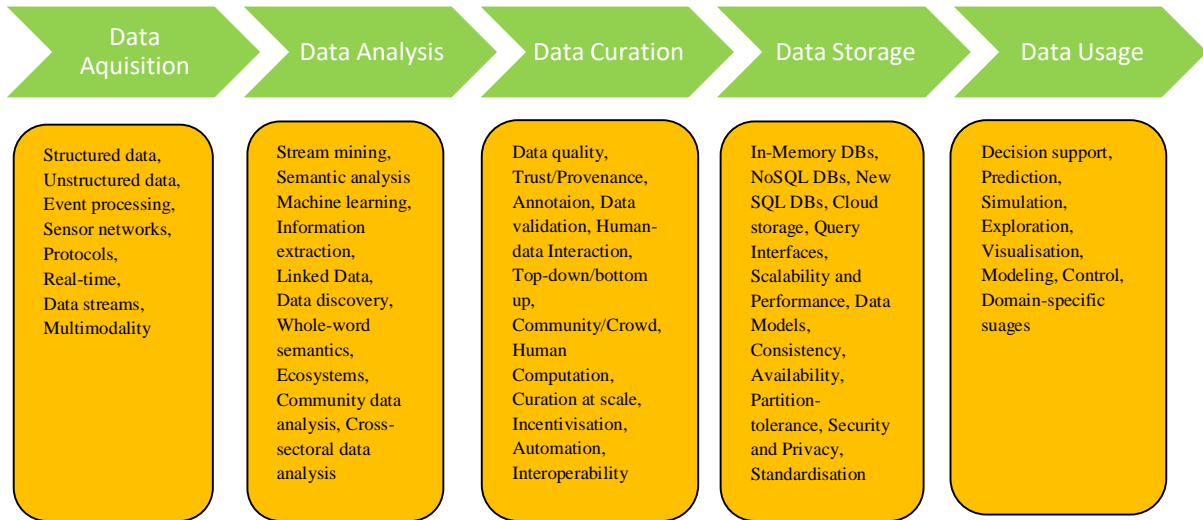
<sup>14</sup> Moore, J. F. (1996). *The death of competition: Leadership and strategy in the age of business ecosystems*. New York: HarperCollins.

- **Technology Providers:** these are typically organisation (large and SME) as providers of tools, platforms, services, and know-how for data management
  - **Data End Users:** Individuals or organisations from different industrial sectors (private and public) that leverage big data technology and services to their advantage.
  - **Data Marketplace:** Individuals or organisations that host data from publishers and offer it to consumers/end users.
  - **Start-ups and Entrepreneurs:** New and emerging small firms that develop innovative data-driven technology, products, and services.
  - **Researchers and Academics:** Investigate new algorithms, technologies, methodologies, business models, and societal aspects needed to advance big data.
  - **Regulators** for data protection and privacy and legal issues.
  - **Standardisation Bodies:** Define technology standards (both official and de facto) to promote the global adoption of big data technology.
  - **NSOs:** National Statistical Offices (NSOs) responsible for the production and dissemination of national official statistics of countries.
- **Investors, Venture Capitalists, and Incubators:** Person or organisation that provides resources and services to develop the commercial potential of the ecosystem.

### 3.1 The Big Data Value Chain

The value chain distinguishes the generic value-adding activities of an organisation allowing them to be understood and optimised. A value chain is made up of a series of subsystems each with inputs, transformation processes, and outputs. Accordingly, as analytical tool, the value chain can be applied to information flows to understand the value creation of data technology. In a data value chain, data/information flows is described as a series of steps needed to generate value and useful insights from data. The Big Data Value as indicated in Figure 4 below can be used to model the high level activities that comprise the data / information system in the ecosystem The Big Data Value Chain identifies the following high-level activities:

**Figure 5: The Big Data Value Chain**



Source: Currey et al. 2014<sup>15</sup>

Policy stakeholders act at the international, regional, national and local level. Looking at the government actors, no visible body emerges as a leading actor to implement innovative data for policy initiatives.

### 3.1 The Stakeholders in the Big Data Ecosystem

The stakeholders in the big data ecosystem range various entities as the big data itself entails many different types. The EU data for policy report<sup>16</sup> distinguishes between the following types of stakeholders: global and European policy makers; national policy makers; regional policy makers; statistical offices; science and R&D organisations; data brokers; private providers of

<sup>15</sup> Curry, E., Ngonga, A., Domingue, J., Freitas, A., Strohbach, M., Becker, T., et al. (2014). D2.2.2. Final version of the technical white paper. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4)

<sup>16</sup> United Nations (2014), p.6. A world that Counts: Mobilizing a Data Revolution for Sustainable Development by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. New York.

data analytics and visualisation tools; civil society and the policy analysis/evaluation community.

### 3.2 The roles of the stakeholders

The roles of stakeholders varies within the framework of the big data ecosystem ranging from producing data to making financial resources available, setting standards and regulatory frameworks, skills and knowledge, brokering/facilitation/capacity building, and providing digital infrastructure and services.

The UN in its global initiative with the launch of ‘A world that counts’<sup>17</sup> initiative states that by 2020, it hopes to be witnessing the emergence of a vibrant ‘global data ecosystem’ to support the monitoring and implementation of the SDGs in which the various stakeholders play their role as stated below:

- **Governments** should empower public institutions to respond to the data revolution and put in place regulatory frameworks that ensure robust data privacy and data protection, and promote the release of data as open data by data producers, and strengthen capacity for continuous data innovation.
- **Multinational organizations, donors, governments and semi-public institution** should invest in data, providing resources to countries and regions where statistical and technical capacity is weak. They should develop infrastructures and implement standards to continuously improve and maintain data quality and usability; keep data open and useable by all. They should also finance analytical research in forward-looking and experimental subjects.
- **International and regional organizations** should work with other stakeholders to set and enforce common standards for data collection, production, anonymization, sharing and use to ensure that new data flows are safely and ethically transformed into global public goods, and maintain a system of quality control and audit for all systems and all data producers and users. They also should support countries in their capacity-building efforts.
- **Statistical systems** should be empowered, resourced and independent, to quickly adapt to the new world of data to collect, process, disseminate and use high-quality, open, disaggregated and geo-coded data, both quantitative and qualitative.
- **All public, private and civil society** data producers should share data and the methods used to process them, according to globally, regionally, or nationally brokered

---

<sup>17</sup> UN (2014). A World That Counts: Mobilising the Data Revolution for Sustainable Development. By the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. UN: November 2014.

agreements and norms. They should publish data, geospatial information and statistics in open formats and with open terms of use, following global common principles and technical standards, to maintain quality and openness and protect privacy.

- **Governments, civil society, academia and the philanthropic sector** should work together to raise awareness of publicly available data, to strengthen the data and statistical literacy (“numeracy”) of citizens, the media, and other “infomediaries”, ensuring that all people have capacity to input into and evaluate the quality of data and use them for their own decisions, as well as to fully participate in initiatives to foster citizenship in the information age.
- **The private sector** should report on its activities using common global standards for integrating data on its economic, environmental and human-rights activities and impacts, building on and strengthening the collaboration already established among institutions that set standards for business reporting.
- **Civil society organizations** and individuals should hold governments and companies accountable using evidence on the impact of their actions, provide feedback to data producers, develop data literacy and help communities and individuals to generate and use data, to ensure accountability and make better decisions for themselves.
- **Academics and scientists** should carry out analyses based on data coming from multiple sources providing long-term perspectives, knowledge and data resources to guide sustainable development at global, regional, national, and local scales. They should make demographic and scientific data as open as possible for public and private use in sustainable development; provide feedback and independent advice and expertise to support accountability and more effective decision-making, and provide leadership in education, outreach, and capacity building efforts.

Therefore, the different stakeholders for big data, which includes owners and users, should ideally emerge into a “global data system”, or big data ecosystem, to support policy making. However, the challenge will be in how to bring these different stakeholders and systems together to make the data revolution happen. These stakeholders are operating within their systems and procedures and it is important that fora and platforms are being established and managed effectively to make the big data system work.

Effective application of Big Data for Development would also require changes in the decision-making process, which customarily relies on traditional statistics. Given the high frequency of Big Data, a more responsive mechanism will need to be put in place that allows the government to process the information and act quickly in response. Also, since Big Data is often unstructured and relatively imprecise (compared to official statistics), government officials also have to learn how to effectively interpret and make use of the information provided by Big Data. This requires capacity building to turn decision makers into more sophisticated data users.

#### **4. Big data and the private sector**

The private sector has increasingly been a major economic contributor to most nations' GDP globally. As a result there is a need for rapid or near-real-time data acquisition within the sector for growth strategies and expansion to reach extended markets with a focus on innovation and cost. The increasing amounts of alternative data gathered and available in the private sector has attracted governments across the board to engage heavily not only with the private sector but also academia, international organisations and CSOs to harness big data for mutual benefit to all.

In this context, the adoption and use of big data by the private sector in Africa can be seen from a couple of different angles. On one hand the use of big data by the private sector for competitive advantage of companies to innovate their business process and enhance productivity in their business domain; and secondly those engaging in the big data analytics business as a startup or following a paradigm shift moving into the big data market

#### **4.1 Participation of the African private sector in the Big Data era**

Across the continent, both home grown businesses and multinational companies are turning to big data to inform their business growth strategies as well as providing other organisations and companies make sense of their customers and businesses. With the proliferation of the tech startups in the continent, data-focused start-ups are also beginning to flourish. The International Data Corporation, a market intelligence firm, estimates that revenue from big data and analytics operations will increase by 11% in Africa and the Middle East this year to reach USD\$2 billion.

The African private sector participation in the global big data market is motivated by the increasing tech hubs in Africa with innovative solutions that made their way to make impact at the global level. These start-ups have proliferated today across the continent and have been dealing with new and emerging technologies that attracted global multinational corporations to look towards the continent for potential market.

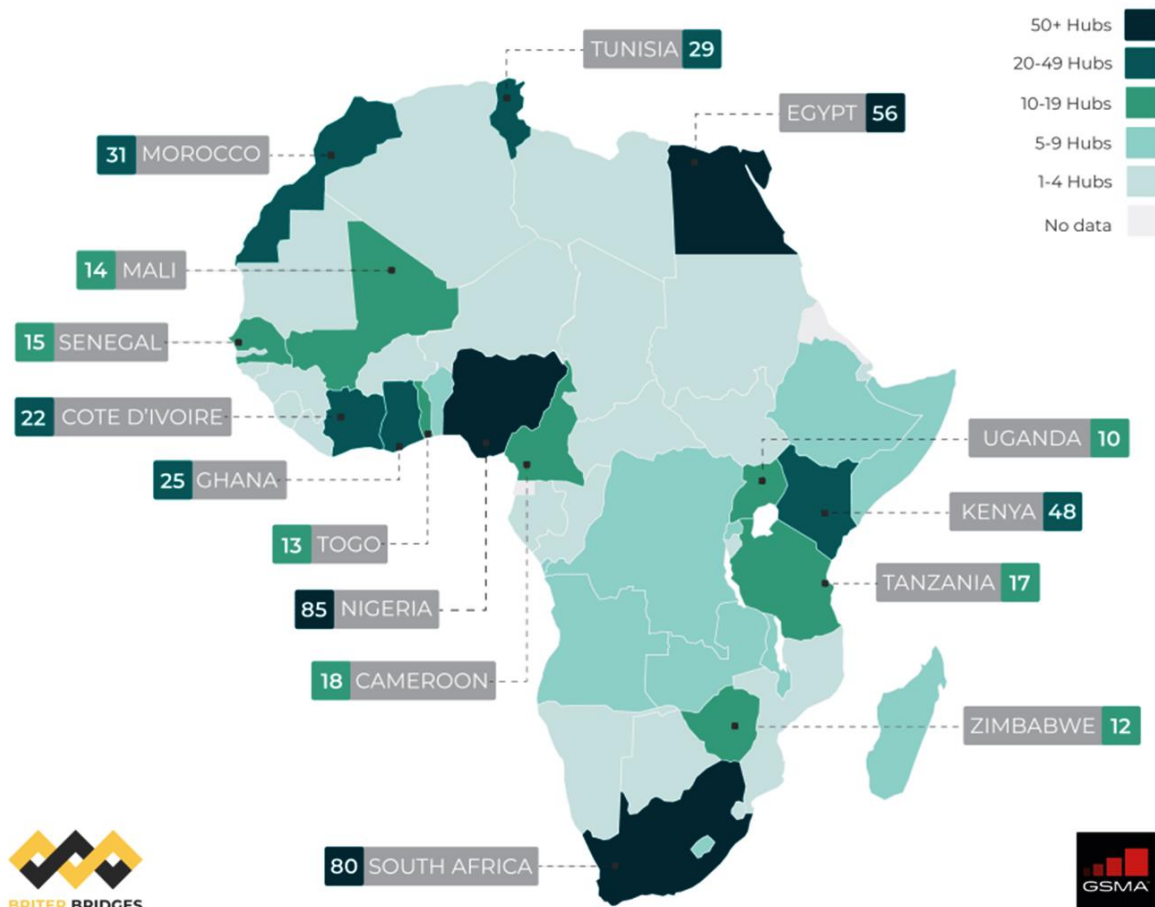
##### **4.1.1 The growth of tech hubs**

The growing tech hubs in Africa have played a great role in the adoption and use of new and emerging technologies ranging from the Internet of Things (IoT) to Big Data, Artificial Intelligence, Blockchain, etc. for various applications ranging from digital finance to e-commerce, and other development applications in the areas of agriculture, water, energy, education, government services, etc.



According to GSMA and Briter Bridges, the tech hubs have reached 618 in 2019 which has jumped from 442 in 2018. The distribution of the tech hubs across the continent is shown in the map below.

**Figure 5: Map of Africa on the distribution of tech hubs**



Source: Briter Bridges and GSMA (2019)

The technological ecosystem in Africa has seen astonishing growth over the last few years, mainly boosted by venture funds, development finance, corporate investment, as well as every expanding innovative community. Accordingly, there are currently 618 active tech hubs, i.e. organisations currently active with a physical address, offering facilities and support for tech and digital entrepreneurs. The tech hubs are distinguished based on the type of support or facility they offer to entrepreneurs, including incubators, accelerators, university-based innovation hubs, maker spaces, technology parks, and co-working spaces (25% of the total active hubs offer co-working). Nigeria, South Africa, Egypt and Kenya are among the top in the number of tech hubs while Nigeria and South Africa remain the most advanced ecosystems

having 85 and 80 active tech hubs respectively. Kenya is also established at the centre of East Africa technology ecosystem with 50 tech hubs, becoming a favoured destination for new investors and corporates to capture the fast-growing pool of tech talents in the country. The growth of the tech landscape in the continent witnesses the rise of new tech cities – established and emerging cities.

#### 4.1.2 The use of big data in development

This tech business landscape together with the expanding Information and Communication Technology (ICT) infrastructure and services provide the enabling environment for the growth of big data uses in various socio-economic sectors in several countries. Some of the use cases with regard to big data applications in selected countries are highlighted below:

One of the early applications of the potential of big data was in the healthcare sector where big data analytics using mobile phone call details to detect and control disease outbreaks as indicated in the two examples below:

- ✓ *Anonymized mobile call-data records (CDRs) are being used to track the movement of people, map the spread of disease in an epidemic and target where treatment centres should be built. And new breakthroughs mean that the tracking can be carried out in real-time. The US Center for Disease Control and Prevention (CDC) is analysing mobile phone mast activity data in Liberia in order to map where calls to helplines are coming from, which indicates where there could be a fresh outbreak of disease. Using real-time data is useful in low income countries where census records are patchy or out-of-date, enabling intervention to be more effective<sup>18</sup>.*
- ✓ *The recent Ebola epidemic in West Africa has provided the testing ground for new technology to track the disease and help prevent its spread. Leading the response was RapidPro, a free open-source software platform hosting apps (IHRIS, mHero and U-Report) developed by UNICEF and the Rwandan software company, Nyaruka. The apps helped central authorities and health workers to communicate quickly and easily with each other, as well as alert health workers on the ground about how to detect signs and avoid the spread of the disease. They can also be used for real-time monitoring. Use of dedicated apps is credited with containing Ebola in Nigeria (the outbreak was limited to a small number of cases) by cutting reporting times and enabling authorities to quickly trace people who may have come into contact with infected patients.*
- ✓ *A Harvard University researcher used records from 15 million mobile phones in Kenya to track the movements of people in an effort to find hotspots where malaria was being transmitted. Analyzing the mobile data, she was able to pinpoint settlements where there*

---

<sup>18</sup> The University of Southampton and the National Vector-borne Diseases Control Programme (NVDCP) have used this technology to help combat the spread of malaria in Namibia.

*was an unexpectedly high volume of parasites and to direct government control efforts to those areas<sup>19</sup>.*

In financial analysis, moving away from the use of non-traditional data to assess credit risks helped smallholder farmers in accessing microfinance facilities as the case below demonstrate:

- ✓ *Kenyan farming app FarmDrive not only allows smallholders to keep a record of their revenues and expenses, it uses that information to assess whether users are eligible for loans. Similarly, Cape Town's Jumo uses mobile data to conduct background checks.*
  
- *The Kenya-based insurance company UAP partnered with Syngenta (providers of famr products) and Safaricom (Kenyan telecom operator) to lauanch the Kilimo Salama (Safe Farming) microinsurance project. Historically, insurance has had many challenges in the Kenyan context, including spam advertising through mobile phones, difficulties claiming insurance money, aggressive, large sales teams and inefficient claims processes. Based on big data from over three decades of climate and crop trends, UAP can determine the appropriate compensation plan for the current year without the need to assess individual cases. This weather index insurance scheme can automatically process insurance claims when the rainfall exceeds an average at a given value. As the first microinsurance product in the world to be fully distributed and implemented over a mobile phone network, farmers can receive insurance policy numbers and premium receipts via short message service (SMS) and insurance payouts via the M-PESA platform. The project was spun off as the company Acre Africa and in 2014 insured a total of 233,796 farmers in Kenya and Rwanda.<sup>20</sup>*

The other interesting use is the potential for start-ups and small businesses to scale up their data efforts, discovering new revenue streams and value adds for customers. Platform-based businesses and start-ups have designed data-centric business model which new ideas and innovation can grow as the example below in Rwanda:

- ✓ *SafeMoto - a Rwandan ride-sharing app that lets people hail motorbike taxis. Its big selling point is that it uses telematics technology to track driver behaviour and kicks those with a low safety score off the platform.*

These two apps are focused on the good use of data to meet unique customer needs. Such initiatives can have a lot of scope to draw the example of Uber Eats type of services and success, through potential strategic partnerships to expand their current initiatives and develop new innovative products for their customers.

The above developments can also show the potential that SMEs can become data-centric in their own business and establish a robust data system in place around their core businesses. Moreover, this is also a potential business opportunity largely untapped among big businesses in the continent including the banking/ insurance companies, telecom operators particularly Mobile Network Operators (MNOs), the retail market, etc. who could have generated additional revenue stream from the existing services and products. This could simply be

---

<sup>19</sup> <http://www.technologyreview.com/lists/innovatorsunder-35/2013/humanitarian/caroline-buckee/>

<sup>20</sup> International Finance Corporation (n.d.); Macharia (2013); <http://acreafrica.com>

achieved through implementing a basic Customer Relationship Management (CRM) system or an open-source business intelligence platform.

### **4.1.3 Business opportunities in Big Data use in government**

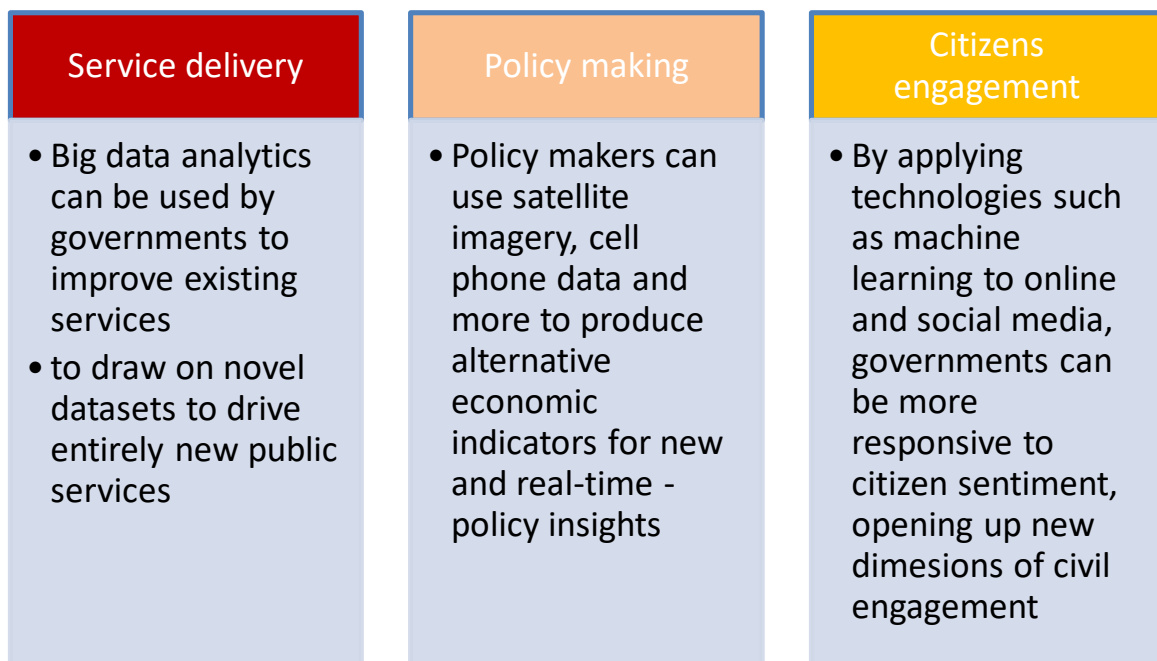
The great amount of data generated, stored, and mined for insights have become economically useful to business, government, and consumers. In the context of policy making, big data can be used to enhance awareness (e.g. capturing population sentiments), understanding (e.g. explaining changes in food prices), and/or forecasting (e.g. predicting human migration patterns). In this regard, in many countries, public sector entities also gather enormous amount of data, for example, from census, tax returns, and public health surveys. Much such data is technically 'public', but accessing is not always easy, and mining it for relevant insights may also require technical expertise and training that organisations and governments with limited resource may not always afford. Therefore, making good use of big data needs collaboration of various actors including data scientists and practitioners, leveraging their strengths to understand the technical possibilities as well as the context within which insights can be practically implemented.

The public sector cannot fully capture the potential of Big Data without active participation and leadership from the private sector. In this regard, there has been discussion about Data Philanthropy – a term which describes a new form of partnership in which private sector companies share data for public benefit, which has been advanced since its emergence of the term at the World Economic Forum in Davos in 2011. The debate on the concept of Data Philanthropy, or private sector data sharing, has been gaining momentum since and moved forward reaching greater audience. In this collaboration, companies and governments need to donate data to help track diseases, avoid economic crises, relieve traffic congestion, and support development. This way, it is not only the government that gain from Data Philanthropy but also private companies could enhance their role in corporate social responsibilities and in return shaping their branding. This development in big data is proving to substantially change, public sector and international development sector in a much more novel way.

Big data in the public sector refers to the use of non-traditional data sources and data innovation to make government solutions more responsive and effective. Governments have opportunities to harness big data solutions to improve productivity, performance and innovation in service delivery and policy making processes. In this regard, government may assume various roles as producer, consumer, and facilitator of big data to enable better service and policy outcomes.

As a producer, government produce millions of datasets ranging from tax returns and unemployment data to hospital funding and energy use. As consumer, governments can use big data both from their own sources as well as from others to promote responsive government. For instance, transport authorities use information from GPS navigation systems to inform traffic management. As facilitator, governments should also invest in big data research and stewardship, including setting up robust legal and technical frameworks to ensure big data is accessible and responsibly used for public good.

Figure 4: Use areas of Big Data in Government



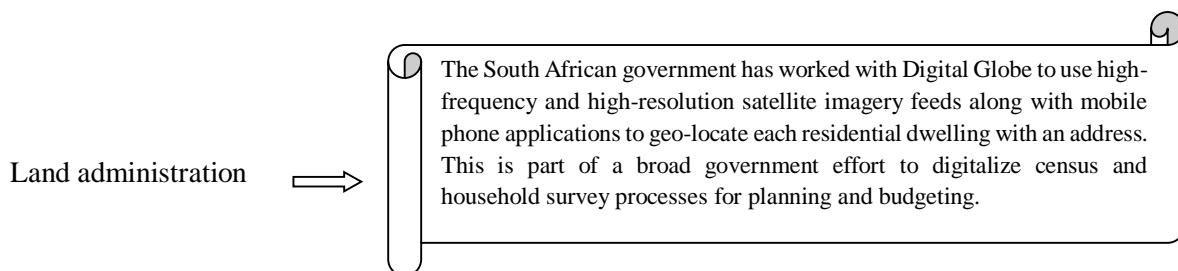
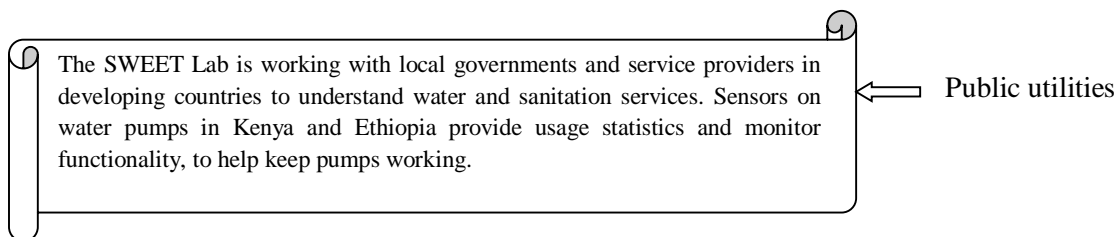
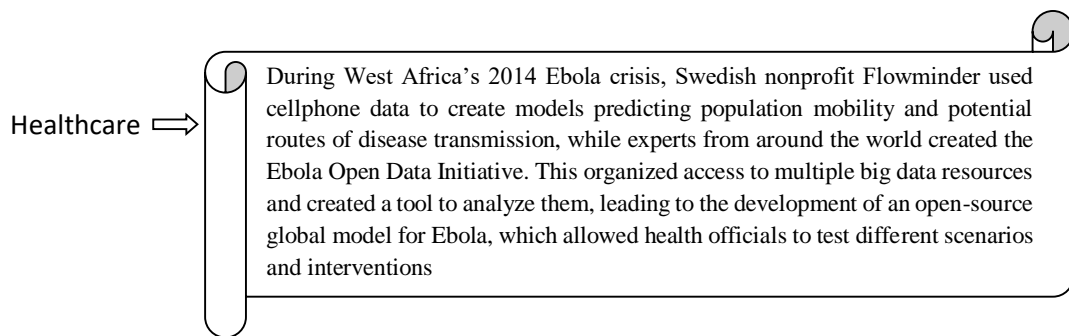
The above figure shows the enormous potential of big data to transform government with applications in service delivery, policymaking and citizen engagement, areas where big data can play a transformational role.

Some of the initiatives that are highlighted below indicate where the government should facilitate action for effectively using data in areas ranging from healthcare to public utilities, transportation, etc<sup>21</sup>.

**Service delivery** – the way the private sector made the customer experience more personalised and consumers’ focused, the society also aspires government services that they interact with

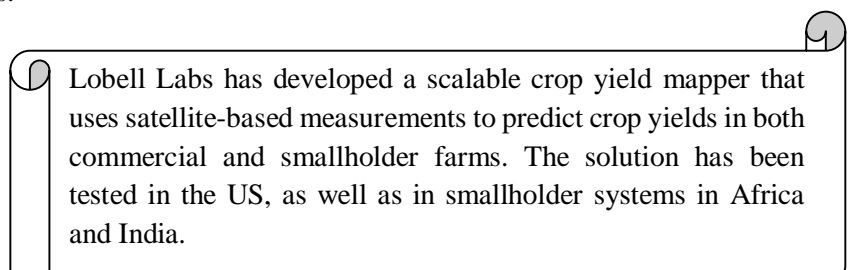
<sup>21</sup> World Bank ( ). Big Data in Action for Government: Big Data Innovation in Public Services. World Bank Group: Washington, DC.

almost on a daily basis whether it is for health, employment, education and business, among others, be big data enabled to make public service delivery equally as smart, responsive and personalised. In this regard, there are several such initiatives on the role of big data in enhancing government service delivery as highlighted below.

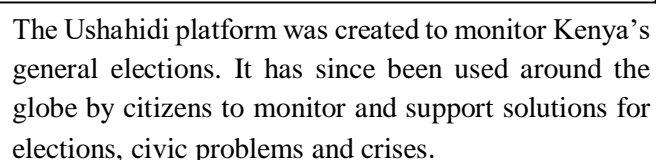


**Policy making** - Big data provides enormous potential as viable source of data to better inform policy decision. Several public-private partnerships are emerging to make data from social media, professional networks, mobile phones and sensors readily available for policy makers.

One such example in the area of food security is the following initiative which also benefits African smallholder farmers.



**Citizen engagement** - the interaction between citizens and government provides enabling environment for improving service delivery and policy making processes. In this regard, big data analytics can enhance this interaction and make much more focused, personalised and responsive. The following example from Kenya demonstrates such citizen engagement which was replicated in various forms across the globe.



The Ushahidi platform was created to monitor Kenya's general elections. It has since been used around the globe by citizens to monitor and support solutions for elections, civic problems and crises.

These are few examples of how the collaboration between the private sector and the public sector helps in harnessing the potential of big data for development goals.

In this regard, McKinsey estimates that open data – public information and shared data from private sources – can help create USD\$3 trillion to USD\$ 5 trillion a year of value in seven areas of the global economy<sup>22</sup>, namely, education, transportation, consumer products, electricity, oil and gas, healthcare and consumer finance.

The African private sector tapped into this market and the World Bank survey identified the open data driven companies in Africa. To this end, in Africa, the Digital Data Divide (DDD) reached out to 41 companies in 11 countries. 34% of these companies were located in Kenya, 27% in South Africa, 10% in Nigeria. Ghana and Tanzania accounted for 7% each. Cameroon, Egypt, Ethiopia, Uganda, Zambia, and Zimbabwe accounted for 15% of the entities.

These companies were primarily active in agriculture, business, health, and telecommunications, though DDD did also hear from companies engaged in

---

<sup>22</sup> Open data: Unlocking innovation and performance with liquid information. Oct 2013 Report. McKinsey Global Institute. By James Manyika, Michael Chui, Diana Farrell, Steve Van Kuiken, Peter Groves, and Elizabeth Almasi Doshi

transportation, real estate, human resources, finance, information technology, environment, insurance, legal, online stores, and security and technology.

Thirty-three companies provided complete or partial financial information. Ghana and Morocco accounted for a third each of the companies willing to provide such information, Kenya for about a fifth, and a couple of companies from South Africa responded as well. The financing needs of the companies surveyed that are trying to raise resources ranged from a few hundred thousand dollars to \$5 million and totalled above \$21,000,000.

## 4.2 Use of big data for competitive advantage of companies

Companies traditionally gather and store data as part of their transactions which they use for tracking or forecasting the future. Now a day, this data has exponentially grown that it requires implementation of new processes, technologies and mechanism called Big Data. Big data has become a billion dollars industry in itself with the global big data and business analytics revenues reaching USD\$ 189.1 billion in 2019. It has three main value chain components, namely, data generation and collection; storage, transmission and security; and analytics. However, it is much more than this and is a source of innovation that generates new products and services. But most importantly, big data is fundamentally changing how companies do business by transforming core business functions and the very nature of competition.

Describing the paradigm-shifting role of big data for businesses, it is stated succinctly that ‘Data now stands on par with people, technology, and capital as core asset of corporations and in many businesses is perhaps becoming the decisive asset’<sup>23</sup>.

In this regard, according to the Mckinsey Global Institute, there are five ways of how big data create value<sup>24</sup>

- It can create transparency by being more widely available to the new potential;
- It enables companies to set up experiments. For example experiments for process changes, they can create and analyse large amounts of data from these experiments to identity possible performance improvements;
- Big data can be used to create a more detailed segmentation of customers to customise actions and prepare specific services. For example, consumer goods and service companies that have used segmentation for many years are beginning to deploy real-time micro-segmentation of customers to target promotions;

---

<sup>23</sup> Porter, Michael E. and Heppelmann, James E., How Smart, Connected Products are Transforming Companies. Harvard Business Review, October 2015 issue. Available at <https://hbr.org/2015/10/how-smart-connected-products-are-transforming-companies>

<sup>24</sup> McKinsey Global Institute (2011). Big Data: The next frontier for innovation, competition, and productivity. Retrieved 20 October, 2019 from [http://bigdatawg.nist.gov/MGI\\_big\\_data\\_full\\_report.pdf](http://bigdatawg.nist.gov/MGI_big_data_full_report.pdf)



- Analysis of Big Data can support human decision making by pointing to hidden correlations or some hidden risks enabled by algorithmic decision support systems. An example can be a risk or fraud analysis engines for insurance companies. Low decision making can be even automated to those engines in some cases. Some organizations are already making better decisions by analysing entire datasets from customers, employees, or even sensors embedded in products.
- Data can also enable new business models, products and services or can improve the existing ones. Data about how products and services are used can be used to develop and improve new versions of the product. For example, the emergence of real-time location data has created an entirely new set of location-based services from navigation to pricing property and insurance based on car driving habits.

By using big data and acknowledging its benefits can companies gain competitive advantage and position themselves ahead of their competitors. Big data provides businesses with much bigger growth potential than traditional technologies. An MIT study found that firms that adopt data-driven decision making have output and productivity<sup>25</sup> that is 5% to 6% higher than what would be expected given their other investments and use of information technology.

Like in other developing countries, the private sector in Africa also values the importance of understanding and analyzing data to drive business outcomes. Nearly half of the business leaders surveyed who were asked what technology they were investing in to better engage with customers or potential customers in emerging markets, they said they were currently investing in data and the analysis of it. Data analytics scored highly with African and Asia Pacific business leaders – 57 per cent and 61 per cent respectively saying that this is an investment area for them. This study found that globally 41 per cent of business leaders said they were significantly increasing spending on data analytics in the next five years to increase their understanding of consumers in emerging market countries<sup>26</sup>.

Telecom service providers, and banks and financial firms have also acknowledged this trend and are embracing the use of data generated by their customers. Around 85% of African banks surveyed by PricewaterhouseCoopers are using big data to improve their security, whereas 77% are using it to improve their customer service.

GSMA estimates the market opportunity for MNOs in IoT Big Data, based on analysis, executive interviews and end-user perspectives and particularly focusing on 2025 as a point at which the estimated market will be worth \$386 billion. Accordingly the breakdown of the various markets within the IoT Big Data market will be a USD\$ 48 billion in cloud and hardware, USD\$110 billion in Big data and analytics software, USD\$ 32 billion in the area of platforms, USD\$78 billion for applications markets, and USD\$118 billion for professional services in the areas<sup>27</sup>.

---

<sup>25</sup> [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1819486](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486)

<sup>26</sup> Long, Jessica and Brindley, William. The role of big data and analytics in the developing world. Accenture and NetHope.

<sup>27</sup> GSMA (2019). The IoT Big Data Revenue Opportunity for Mobile Operators. Prepared by PwC: October 2019

This shows a promising market potential in the continent that the private sector should position and develop strategies to capture this new exciting marketplace. Furthermore, business leaders recognise the huge potential to strengthen customer engagement and insight. South African companies have been pioneering the use of sensors to monitor driving as they look to reduce accidents. They are also leading the way in the use of health monitors to detect signs of ill health and remind people to take regular medicines.

## **5. Challenges**

There are many challenges with big data that should be considered to foster the private sector in Africa. The discussion of big data is quite complex, ranging from practical or technical challenges to legal and regulatory limitation. As one of the outcomes of this report being providing rigorous analysis and peer learning, translating findings into policy dialogue and recommendations. It focuses on big data in the context of fostering the private sector towards this age of big data world. Some of the This report identifies on some of these challenges and doesn't pretend to provide answers and solutions but rather to promote discussion.

### **Access to Big Data**

As seen in this report above, African private sector use of open data available from the public sector has been a source of innovation and business opportunity. As a producer of data, government produce millions of datasets ranging from tax returns and unemployment data to hospital funding and energy use. Access to this data should be facilitated by the government and made openly available which in many countries is not the case.

### **Privacy, confidentiality and cybersecurity risks**

As Big data involves large amounts of sensitive and personal information that could be exposed to privacy, confidentiality and cybersecurity risks. Such data, if not properly handled is vulnerable to cyberattacks, used to profile individuals or organisations, and sold to third parties. Although data / information may not be used legally without a person's prior permission or knowledge, depending on the national legal framework, the unauthorised use and loss of data could render reputational damages as well as loss of consumers' trust.

### **Legal frameworks on data protection and privacy**

Not many countries have put in place the necessary legal and regulatory frameworks in ensuring the data protection and privacy in the use of the cyberspace. However, data protection and privacy legal frameworks, procedures and technological infrastructure to protect such rights will be key to minimising privacy, confidentiality and cyber-security risks when using big data.

### **Financial challenges**

For start-ups and other business to move into the big data business, the challenge has been the cost of investment although now tools such as cloud technologies may relieve some of the burden of the cost of investment. Most of the innovative initiatives have been supported by international development funding. In recent times, Africa has also gained popularity in attracting international multinational corporations' attention in investment and support to start-ups which should be supported by creating an enabling environment on the part of the government. Therefore, although well funded many businesses and other national development and research systems are adapting rapidly to capture the data challenge, the great potential of big data remain remote for many countries especially where ICTs are less prevalent where the costs of adaptation pose particular problems..

### **Building a critical mass of technical skills**

The UN Global Working Group concluded that a multidisciplinary team from different professional backgrounds is necessary to adapt the analytics to the data. Therefore, the private sector need to develop such multidisciplinary skills in addition to building a team familiar with big data – for example, data scientists, IT architecture specialists, and data visualisation specialists who work alongside subject matter professionals. Furthermore, there is no network established that allows businesses to tap into expert with specific skills.

### **Infrastructure and services**

Although Africa is much better in terms of infrastructure and services compared to a decade earlier, there is still a long way to go to make a competitive environment for the vibrant innovators and the booming start-ups in the continent. For example, cloud technology offers for the African market a flexibility to scale the technical infrastructure up or down depending on the needs. However, the preferred approach in the industry and typically can require a budgetary allocation of about USD\$ 250,000 for the first year in which subsequent years are based on 'pay for what you use' and efficient use for optimum business can be planned. Therefore, African countries need to develop and provide the necessary infrastructure and service to grow the innovative business community.

## **6. Conclusion and recommendations**

## **6.1 Conclusions**

Disruptive technology is transforming Africa's economic potential. This includes creating millions of new targetable consumers and giving them unprecedented choice of products and services.

It's also fostering the innovation and connectivity that can create new markets and new business models. Indeed, because Africa is not dependent much on long years of legacy technology, it is able to leapfrog more than developed economies. And combined with demographic changes and urbanisation, digital disruption means big opportunities for businesses and investors across the continent. It is therefore important for the continent to seize this opportunity to transform Africa's digital revolution into creating a vibrant digital economy.

To realise these opportunities, both the private sector and governments need to be ready to face the challenges of the emerging disruptive technologies that still lie ahead. Among the key priority to this effect include developing the skills, the telecom infrastructure and services, the transport links, the energy supply, etc. that will enable digital disruption has the infrastructure and capacity to be genuinely transformational. Governments have to equally ensure that the reach and benefits of digital technology and the growth it brings are evenly distributed as the trend shows that the upcoming tech hubs across the continent move also away from the capitals and reaching out emerging cities in the continent.

## **6.1 Recommendations**

The following are some recommendations identified from the findings of this report for possible considerations by businesses and policy- and decision-makers:

- African mainstream businesses need to develop agile and adaptable business model for embracing big data to work in different African markets.
- Government and other stakeholders need to support upcoming yond start-ups to further investment in innovative products to scale-up and move beyond start-up and becoming a leading market player. Start-ups also need to network and use their data and customer relationships to develop new products and services to create new revenue opportunities.

- African governments need to capture the booming tech hubs to create or change regulations to support growth of entrepreneurialism;
- African governments need to ensure and make the necessary reform in the education system to generate the necessary people with key technical skills for the big data ecosystem
- African governments need to adapt to the emerging technologies and the digital disruption environment to improve policy making and deploy recourse more efficiently
- Countries need to elaborate the interface between big data and the new phenomenon of ‘open data’ to promote innovation and to make the potential of big data much more powerful and useful.

## **Annex I**

### **Big Data Types**

Variety is one of the principles of Big Data as described previously. The Big Data can be divided into three types<sup>28</sup>: Structured Data, Semi-Structured Data, and Unstructured Data. Definitions and examples of each can be described as follows:

#### **Structured Data**

---

<sup>28</sup> Maaroo, Abbas. Big Data and the 2030 Agenda for Sustainable Development. UNECE

Structured data generally refers to data that has a defined length and format. Most organizations are storing large amounts of structured data in various divisions, in normalised/ deformed formats in a database: Data warehouses, relational database management system (RDMSs), and various other environments. The data can be queried using a language like structured query language (SQL) in which the datasets can be updated with new data, and deleted, read or any other activity.

The evolution of technology provides newer sources of structured data being produced - often in real time and in large volumes. The sources of data are divided into three categories:

***(a) Computer- or Machine-Generated Structured Data***

Machine-generated data generally refers to data that is created by a machine without human intervention. They can include the following:

- **Sensor data:** Examples include radio frequency ID (RFID) tags, smart meters, medical devices, and Global Positioning System (GPS) data. Another example of sensor data is smartphones that contain sensors like GPS that can be used to understand customer behavior in new ways. For example, RFID is rapidly becoming a popular technology. It uses tiny computer chips to track items at a distance. An example of this is tracking containers of produce from one location to another. When information is transmitted from the receiver, it can go into a server and then be analyzed. Companies, for example, are interested in this for supply chain management and inventory control.
- **Web log data:** When servers, applications, networks, etc operate, they capture all kinds of data about their activity. This can amount to huge volumes of data that can be useful, for example, to deal with service-level agreements or to predict security breaches.
- **Point-of-sale data:** When the cashier swipes the bar code of any product that you are purchasing, all that data associated with the product is generated. Just think of all the products across all the people who purchase them, and you can understand how big this data set can be.
- **Financial data:** Lots of financial systems are now programmatic; they are operated based on predefined rules that automate processes. Stocktrading data is a good example of this. It contains structured data such as the company symbol and dollar value. Some of this data is machine generated, and some is human generated.

***(b) Human-Generated Data:***

This is data that humans, in interaction with computers, supply.

- **Input data:** This is any piece of data that a human might input into a computer, such as name, age, income, non-free-form survey responses, and so on. This data can be useful to understand basic customer behavior.
- **Click-stream data:** Data is generated every time you click a link on a website. This data can be analyzed to determine customer behavior and buying patterns.
- **Gaming-related data:** Every move you make in a game can be recorded. This can be useful in understanding how end users move through a gaming portfolio.

The way data is structured is a vital element. If the structures aren't coherent and understandable, data is liable to be misused (misunderstood) and will fail to facilitate "bringing together" data from disparate sources to produce new knowledge/evidence.

This is a metadata schema related issue – or brought down to a simple example what headings/terms are being used for columns of data in a spread sheet and how can the person using the spread sheet understand the context.

### **Semi-Structured Data**

Semi-structured data is a kind of data that falls between structured and unstructured data. This type of data became a talking point. Mostly data coming from Facebook, Twitter, Blogs, publically available websites, etc. makes the basis of semi-structured data. These data sources usually have defined structures and mostly contain text information.

The free flow text generated through the social media is the only unstructured component whilst the remaining data is structured. Most of the times, the social data is mistaken with unstructured data. The social data is NOT unstructured data, it is semistructured and in fact, some of the social data contains industry standard structures.

- **Social media data:** This data is generated from the social media platforms such as YouTube, Facebook, Twitter, LinkedIn, and Flickr.

### **Unstructured Data**

DataUnstructured data does not have any defined, consistent fields and it may even do not have any numbers and text. Unstructured data can be divided also into either machine generated or human generated and described as flows:

#### ***(a) Machine-Generated Unstructured Data Examples***

- **Satellite images:** This includes weather data or the data that the government captures in its satellite surveillance imagery. Just think about Google Earth, and you get the picture (pun intended).
- **Scientific data:** This includes seismic imagery, atmospheric data, and high energy physics.
- **Photographs and video:** This includes security, surveillance, and traffic video.
- **Radar or sonar data:** This includes vehicular, meteorological, and oceanographic seismic profiles.

#### ***(b) Human-generated Unstructured Data Examples***

- **Mobile and Voice data:** This includes data such as text messages and location information. Human voice contains a lot of information and it needs access and mined. The spectrogram of the human voice reveals its rich harmonic content including pitch, tone, emotion, bass, etc.
- **Web behavior and content:** This comes from any site delivering unstructured content, like YouTube, Flickr, or Instagram. The scope of web behavior is huge. There are nearly five billion indexed web pages on the Internet and for each page there are traffic statistics ranging from the number and duration of visits to far richer information on

user behavior on a large proportion of websites. Big Data also encompasses the content of those web pages and the changes that occur on them. Also included in this category is the vast amount of search engine data constantly being generated.

- **Image and Video Data:** Total number of pictures taken in last 5 years is more than double the pictures taken in 1900 -2000. This gives us an opportunity to use patterns within the pictures and mine the information available to us. Various techniques like pixilation, pattern matching, image processing, feature extracting, etc. allows us covert the pictures into data and further mine it using classification algorithms. Examples of image data use cases: One of the most

## **Annex II**

Selected companies in Africa engaged in open data based business



Company Name	Country	City	Website
Mobile Content.Com Ltd.	Ghana	Accra	<a href="http://www.mobilecontent.com.gh">http://www.mobilecontent.com.gh</a>
Etranzact Ghana LTD	Ghana	Accra	<a href="http://www.etranzactgh.com">http://www.etranzactgh.com</a>
Universal Hospitals Group Ltd	Ghana	Accra	<a href="http://www.uhgl-gh.com">http://www.uhgl-gh.com</a>
E-Transactions Ghana Ltd	Ghana	Accra	<a href="http://www.e-transactions.org/index.php">http://www.e-transactions.org/index.php</a>
Aspect Foundry	Ghana	Accra	<a href="http://www.aspectfoundry.com">http://www.aspectfoundry.com</a>
Syecomp Business Services Ltd	Ghana	Accra	<a href="http://www.agricinghana.com">http://www.agricinghana.com</a>
RLG GHANA	Ghana	Accra	<a href="http://www.rlgglobal.com">http://www.rlgglobal.com</a>
Oxford Research Academy	Ghana	Accra	<a href="http://www.oxfordresearchacademy.com">http://www.oxfordresearchacademy.com</a>
JCS Investments	Ghana	Accra	<a href="http://www.jcs.com.gh">http://www.jcs.com.gh</a>
AfroChic Limited	Ghana	Accra	<a href="http://www.afrochiconline.com">http://www.afrochiconline.com</a>
Greenish Bright Farms Ltd	Ghana	Dormaa	<a href="http://www.greenishbright.com">http://www.greenishbright.com</a>
Agro Cocoa Investment Limited	Ghana	Kumasi	<a href="http://www.agrococoainvestment.com">http://www.agrococoainvestment.com</a>
Edu Live	Kenya	Nairobi	<a href="http://elearn-africa.org">http://elearn-africa.org</a>
Ufahamu Kenya	Kenya	Nairobi	<a href="http://www.ufahamu.or.ke">http://www.ufahamu.or.ke</a>
MA3Route	Kenya	Nairobi	<a href="http://www.ma3route.com">http://www.ma3route.com</a>
Majisoft	Kenya	Nairobi	<a href="http://synacor.co.ke/">http://synacor.co.ke/</a>
Mkulima Young	Kenya	Nairobi	<a href="http://www.mkulimayoung.co.ke/">http://www.mkulimayoung.co.ke/</a>
Zege Technologies Ltd.	Kenya	Nairobi	<a href="http://www.zegetech.com">http://www.zegetech.com</a>
African Laughter	Kenya	Nairobi	<a href="http://www.africanlaughter.com">http://www.africanlaughter.com</a>

Manaliagri	Morocco	Agadir	
Egora	Morocco	Agadir	
Extraflore	Morocco	Agadir	<a href="http://www.extraflore.com">http://www.extraflore.com</a>
B&Tech intelligence	Morocco	Casablanca	<a href="http://www.b-techs.org">http://www.b-techs.org</a>
MidwaycoSolar	Morocco	Fes	<a href="http://www.midwaycosolar.com">http://www.midwaycosolar.com</a>
Africmaroc	Morocco	Rabat	<a href="http://www.ambatiment.ma">http://www.ambatiment.ma</a>
B3G Technologies	Morocco	Rabat	<a href="http://www.b3gtech.com">http://www.b3gtech.com</a>
Caciopee	Morocco	Rabat	<a href="http://www.caciopee.biz">http://www.caciopee.biz</a>
CST	Morocco	Rabat	<a href="http://www.cst.ma">http://www.cst.ma</a>
Geospheres	Morocco	Rabat	<a href="http://www.geosphereservices.com">http://www.geosphereservices.com</a>
Matekap	Morocco	Rabat	<a href="http://www.matekap.com">http://www.matekap.com</a>
NavCities	Morocco	Rabat	<a href="http://www.navcities.com">http://www.navcities.com</a>
Naija Connect Store	Nigeria	Kaduna	<a href="http://www.naijaconnectstore.com.ng">http://www.naijaconnectstore.com.ng</a>
Korbitec	South Africa	Cape Town	<a href="http://www.korbitec.com">http://www.korbitec.com</a>
Korwe Software	South Africa	Cape Town	<a href="http://www.korwe.com">http://www.korwe.com</a>
Last Mile for BoP	South Africa	Cape Town	<a href="http://www.lastmileforBoP.com">http://www.lastmileforBoP.com</a>
Pricecheck	South Africa	Cape Town	<a href="http://www.pricecheck.co.za/">http://www.pricecheck.co.za/</a>
Twenty Fingers	South Africa	Cape Town	<a href="http://www.filmcontact.com/member/stefan-coetzer">http://www.filmcontact.com/member/stefan-coetzer</a>
Accura Analytics Systems	South Africa	Pretoria	<a href="https://www.facebook.com/AccuraAnalyticsSystems">https://www.facebook.com/AccuraAnalyticsSystems</a>
Africa Research	South Africa	Pretoria	<a href="http://www.africaresearch.co.za">http://www.africaresearch.co.za</a>
TMS Research	South Africa	Sandton	<a href="http://www.tmsresearch.co.za">http://www.tmsresearch.co.za</a>
Journey Mobile Pty Ltd	South Africa	Stellenbosch	<a href="http://www.journeyapps.com">http://www.journeyapps.com</a>
Problem Solved Ltd	Tanzania	Dar Es Salaam	<a href="https://vc4africa.biz/members/ericmutta/profile/">https://vc4africa.biz/members/ericmutta/profile/</a>
Farming instructor	Tanzania	Dodoma	<a href="http://mvi gour.com/FarmingInstructor.html">http://mvi gour.com/FarmingInstructor.html</a>

