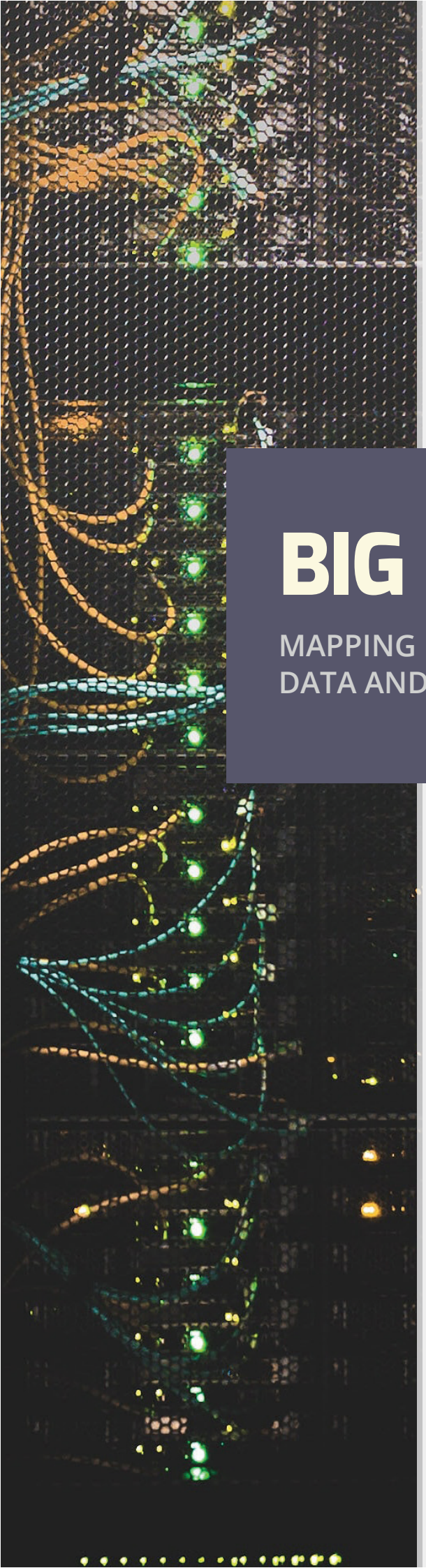


BIG DATA FOR DEVELOPMENT

BIG DATA IN AFRICA

MAPPING EMERGING USE CASES OF BIG DATA AND RELATED TECHNOLOGIES

LEONIDA MUTUKU

LDRI local development research institute

AI:AI Accountable Inclusive Artificial Intelligence

BIG DATA IN AFRICA

Mapping Emerging Use Cases Of Big Data And Related Technologies

LEONIDA MUTUKU

April 2020

LOCAL DEVELOPMENT RESEARCH INSTITUTE

Images courtesy of users Paul Kagame on Flickr and Taylor Vick and the National Cancer Institute on unsplash.com

This work was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada, www.idrc.ca. and with financial support from the Government of Canada, provided through Global Affairs Canada (GAC), www.international.gc.ca

We are an action oriented think tank that supports efforts of African countries to end hunger, extreme poverty and reduce inequality.

Our works examines the state of evidence used for decision making including the state of the institutions producing, the evidence and those relying on that evidence in order to deliver on their mandates.

We have expertise and experience in public policy, open data, data science, artificial intelligence and capacity building of public sector institutions and actors.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The State of Big Data for Development in Africa is a research project designed to map emerging use cases of big data and related technologies being deployed on the African continent towards achieving the goal of sustainable development.

This is with the aim of understanding existing gaps and how the BD4D Africa network can support initiatives that leverage big data to accelerate their achievement of the continent's development priorities. LDRI conducted a landscape study that offered a critical review on the state of big data in Africa, particularly towards addressing challenges related to the social and economic development of African countries. The study mainly highlighted the following findings:

Big Data efforts are mainly driven by private organizations (largely big internet and tech companies, international organizations and large multinationals), research organizations and universities. This is primarily due to existing capacity and access to infrastructure within these institutions to conduct analysis with big data. However, governments still require a lot of support, capacity building and partnerships in order to mainstream the use of big data to achieve their development priorities.

While most projects involving big data do have the potential to accelerate development, they are primarily conducted for private gain and incentives are therefore required to encourage institutions to make more investments that directly target achievement of SDGs (e.g. opening up big data for use by other actors). A lot of development initiatives leveraging big data are conducted in the form of partnerships between various ecosystem actors. However, several of these do not proceed beyond the pilot stage, primarily due to limitations to funding as well as the problems that emerge when governing multiple stakeholders with various motivations.

There are very few institutions across the continent training new big data experts and those that are have mostly partnered with private organizations such as Facebook, Oracle, IBM and

Google to deliver big data training. There is a need for increased investment in capacity building and education in this space to ensure that local contexts, entrenched biases and nuances are considered even as we promote collaboration with international partners.

There are weak regulatory frameworks that both enhance or protect the usage of big data (especially personal data) in big data projects in development. As alternative sources of personally identifiable data are widely accepted as sources to be harnessed in development planning and implementation, LDRI advocates for African countries to ensure the right protections for privacy and consent are in place, and the right balances between innovation and human rights are struck. These recommendations provide a foundational framework for future network activities and programs, especially for the Africa hub of the BD4D network.

INTRODUCTION

The State of Big Data for Development in Africa is a research project designed to map emerging use cases of big data and related technologies being deployed on the African continent towards achieving the goal of sustainable development.

Big data applications have the potential to catalyze the solution of local development challenges in each African country, including applications in key sectors of the economy – agriculture, energy, resource management etc. By leveraging on big data, development partners can quantify inefficiencies in systems, resource allocation needs, and how best to provide interventions that impact every citizen.

LDRI conducted a landscape study and offers a critical review on the state of big data in Africa, particularly towards addressing challenges related to the social and economic development of African countries. The study aims to offer foundational insights in the creation of appropriate programs and activities within the Africa Hub of the Big Data for Development Network.

Therefore, in order to meet this objective, the study aimed to answer the following key questions:

a. **INFRASTRUCTURE**: What are the existing and potential sources of big data available in Africa? What infrastructure and technologies are used in the production and usage of big data?

b. **ECOSYSTEM**: Who are the key actors, ecosystems and initiatives producing and/or using big data? What are their capacities, competencies and short-comings? How are

their activities or initiatives impacting the progress towards achieving set development outcomes?

c. **CONTEXT**: What models for engagement and collaboration, management mechanisms and other contextual considerations need to be made to ensure that the continent fully harnesses the potential of big data in addressing its development challenges?

d. **PARTNERSHIP**: In what ways can the African Hub offer support to existing and potential stakeholders aiming to use big data in the context of development? What activities, additional research, access to tools and infrastructure, network and partnership development, and strategic programming can be designed to better support these stakeholders?

In order to effectively carry out this research, LDRI conducted a review of literature, reports and other documentation emerging from different scholarly publications, news articles and other online material to answer 'Who, What, Why, Where and How'. This review formed our understanding of any existing ecosystems, impact and progression of big data for development projects, and implementation challenges encountered by institutions implementing big data technologies. As part of the study, we also analyzed the policies and other legal frameworks ratified that promote the use of big and alternative data to address development challenges.

BIG DATA IN THE CONTEXT OF DEVELOPMENT

DEFINING BIG DATA

The concept of big data was first coined in 2005 by O'Reilly Media's Roger Mougalas, although its usage is recorded further back in history where it was used to understand phenomena such as disease outbreaks, agriculture and trade records[1]. Data is indispensable in today's administrative operations, business and technology world. Big data can be described as large volumes of data in both structured and unstructured formats used by organizations on a day-to-day basis. Although there are various definitions of big data to expound on this concept, the following three components to big data scope it efficiently:

» **Size**: big data comprises large datasets, where terabytes or petabytes of storage is required.

» **Complexity**: big data is usually in the form of complex, often unstructured datasets, which cannot be analyzed immediately using traditional methods.

» **Technologies**: big data requires powerful tools and special techniques to be processed, analyzed and represented due to its complexity and size[2].

Further, 4 V's of big data - volume, velocity, variety and veracity – describe its unique properties due to the accelerated pace of production of the data in today's digital world and an increasing number of diverse sources that one can tap into to describe a phenomenon.

Big Data is evolving at an exponential rate and the cost of analyzing it has decreased at a corresponding rate as innovations have improved data processing speeds and reduced cost of access to big data storage and analysis tools. This provides a compelling case for governments

1 Rijmenam, M. A Short History of Big Data https://datafloq.com/read/big-data-history/239

2 Ward, Jonathan Stuart, and Adam Barker. 2013. "Undefined By Data: A Survey of Big Data Definitions." Arxiv.org.

and public institutions to utilize big data strategically to enhance service delivery by optimizing usage of public resources, waste and cost reduction in the planning and budgeting processes and deeper citizen engagement.

Every institution including government operations, relies heavily on data for its planning, coordination, implementation and more so, sustainability of public service delivery projects to its citizens whilst also creating an enabling environment for the growth of businesses and enterprise opportunities. According to the Secretary-General's Independent Expert Advisory Group of the United Nations (IEAG), countries are now making significant efforts to incorporate big data analytics in the race to improve statistical capacity for evidence-driven decision-making on food insecurity, spread of diseases, transport and policy implementation among other development aspects[3].

The 17 Sustainable Development Goals (SDGs) are ambitious in that they describe 169 multi-dimensional targets to be achieved by 2030. In total there are 232 statistical indicators that guide each country in their measurement of their performance towards achieving the SDG targets[4]. Countries therefore need to turn to big data in order to not only monitor their progress towards achieving their development goals, but also as resources for innovation - to build data-driven products that they can utilize to achieve the same development goals.

THE AFRICAN DATA REVOLUTION

The United Nations IEAG on the Data Revolution for Sustainable Development defines the data revolution in the context of sustainable development as:

» The integration of these new data with traditional data to produce high-quality information that is more detailed, timely and relevant for many purposes and users, especially to foster and monitor sustainable development;

» The increase in the usefulness of data through a much greater degree of openness and transparency, avoiding invasion of privacy and abuse of human rights from misuse of data

3 'A WORLD THAT COUNTS' Presented to Secretary-General, Data Revolution Group https://www.undatarevolution.org/report/

4 Big Data and The Sustainable Development Goals, The Organization of World Peace https://theowp.org/big-data-and-the-sustainable-development-goals/

on individuals and groups, and minimizing inequality in production, access to and use of data;

Ultimately more empowered people, better policies, better decisions and greater participation and accountability, leading to better outcomes for people and the planet".

» The 2016 edition of the African Data Revolution Report went a step further to define the African data revolution as one which involves:

» The exponential increase in the volume, types and speed of data available in African countries;

» The increased availability and use of new types of data as well as new uses of old or conventional types of data;

» New principles of data governance, including the principles of inclusion and openness; and

» New data-related rights and freedom[5]

This scenario, the report further argued, is only possible with a fundamental conceptual and paradigmatic shift on the question of who and what officially counts, is counted, how, by whom, for whom and for what purposes.

African more generally ascribes to a unified Agenda 2063, beyond the normative and development frameworks developed internally in each country. As such, applications of big data for development are not just important to drive global policies but also to support each individual country in their pursuit of their own individual development plans. They, therefore, have to be designed to bear in mind local contexts, social, economic and regulatory structures.

That is, the big data revolution has to be localized.

SOURCES OF BIG DATA

Traditionally, national statistics offices (NSOs) have collected data from the population through national surveys. However, with the introduction of digital tools, there is an opportunity to improve the speed, accuracy and efficiency in data collection for planning and development purposes. The data revolution however argues for diversification of data sources - moving

5 ECA, UNDP, Open Data for Development Network, World Wide Web Foundation. 2016. "The Africa Data Revolution Report." https://www.uneca.org/sites/default/files/uploaded-documents/ACS/africa-data-revolution-report-2016.pdf

beyond the current statistics produced by NSOs, to also include new sources of data created by digital and mobile platforms.

Africa is increasingly becoming digital with over 44% mobile subscription rates and 23% mobile internet penetration rates[6]. Some countries like Kenya have already achieved 100% subscriber penetration. Further, widespread adoption of mobile technology and consequently the use of social media channels such as Facebook, Twitter and WhatsApp on mobile phones, has not only made the use of digital platforms the norm but also given rise to yet another revolution, the 'mobile revolution'[7].

These platforms carried on mobile devices create what is known as data exhaust, the meta-data generated as a by-product of people's online actions. When mined, this data exhaust reveals a lot of new information about consumers, their behaviours and their livelihoods, as mobile phones become part and parcel of everyday life. This type of data is what makes up the majority of big data available.

Consequently, data mined from mobile phones is touted to be at the heart of Africa's big-data phenomenon. Mobile data provides the granularity that development partners wish could come from census data – it is available in near-real time. Insights from such mobile data can be derived for instance, from migration patterns and used for a variety of use cases such as understanding associated public health concerns, trade and economic implications to financial inclusion, sectors that are key tenets for sustainable development. This kind of data starts to

6 GSMA Intelligence. 2020. "The Mobile Economy Sub-Saharan Africa 2019 ." https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/03/GSMA_MobileEconomy2020_SSA_Eng.pdf.

7 Anderson, Mark The hunt begins for Africa's big data. The Africa Report https://www.theafricareport.com/791/the-hunt-begins-for-africas-big-data/

become very critical for developing countries that do not have sophisticated national statistics collection methods or usable data for critical monitoring and development reporting.

The African data revolution however, goes beyond mobile and incorporates various new sources such as:

» Open data from governments – there are now more than 17 open government data portals hosting important datasets related to budgets, education, agriculture and healthcare[8].

» Citizen-generated data or crowd-sourced data collected on big data platforms such as Ushahidi and repositories such as Zindi Africa

» Sensors data and devices such as those used by Aerotobics to monitor agriculture activities

» Digital transaction systems such as – E-commerce platforms, ERPs and POS, of which there are several local versions in each country.

» Satellite imagery data e.g. GPSDD's Africa Regional Data Cube and Digital Earth Africa's Open Data Cubes used to understand urban infrastructure development and land use

» Electronic medical records to track patient histories and disease prevalence

» Astronomy projects such as The Square Kilometre Array (SKA) project, an international collaboration to build the world's largest radio telescope co-located in Africa and in Australia, by 2024.

8 2019. "The State of Open Data: Histories and Horizons." Open Data for Development. https://www.stateofopendata.od4d.net/.

Members of the Big Data For Development Network host a session at the 2018 Word Data Forum in Dubai.

THE BIG DATA FOR DEVELOPMENT ECOSYSTEM

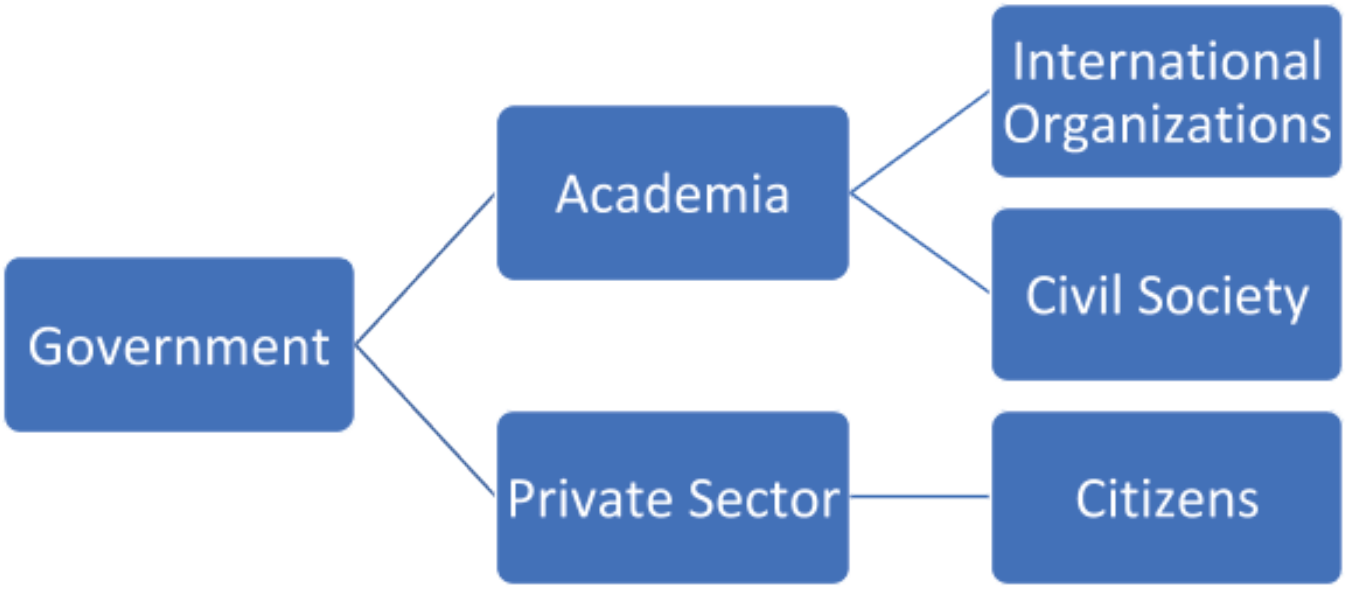
According to the Open Data Charter, a data ecosystem is the complex system of relationships between individuals, organizations, data sets, standards, resources, platforms and other elements that define the environment in which each particular data resource exists.

Expectedly, the big data for development ecosystem comprises different actors with different motivations and playing complex roles to implement big data initiatives. Each of these actors contribute to the production of the big datasets, the development of data infrastructure and tools, the regulatory frameworks and the funding for big data initiatives.

Government, through its ministries, parastatals and agencies tends to plan, implement and, in most cases, fund big data initiatives in their countries. National Statistics Offices are especially central to governments' big data strategies. However, as earlier mentioned, NSOs are required to evolve to meet the demand from government stakeholders who are implementing big data.

It is observed that many African Governments intending to utilize big data in advancing development, are heavily reliant on expertise, funding and technology from multilateral organizations to carry out these exercises. This has greatly shifted how they identify new problems, improve capacity to understand and use data effectively to improve the lives of their citizens and increase business productivity. Increasingly, in the context of the data revolution, strong ecosystem relationships between government and private actors are becoming the norm. Most initiatives that are rolled out by government comprise private-public partnerships or are advised by a diverse taskforce representing diverse communities.

Private sector has over the years increasingly become the prevalent economic contributor to most nations' GDP globally. This has clearly necessitated the need for rapid or near-real-time data acquisition for growth strategies and expansion to reach extended markets with a key



focus on innovation and cost. This availability of vast amounts of alternative data collected in the private sector has thus propelled governments across the board to engage heavily not only with the private sector but also academia, international organizations and civil society organizations (CSO) to harness big data for the mutual benefit of all parties. In most cases, private sector stakeholders are responsible for sharing the big data they collect, in accordance with regulatory frameworks in place.

In many cases, retail, communications and financial services sectors seem to be the early adopters of big data as they have also been the industries that have faced rapid disruption in the last decade. Their adoption of big data therefore has been in a bid to reduce operational costs, improve customer experience and satisfaction and generate revenue from new business models. This data used by these companies is however usually heavily regulated and not easily available in open domains.

Large technology and Internet companies such as IBM, Amazon, Microsoft, Google, and Facebook have leveraged their capacity and created powerful tools that are not only for business gains but also generate data exhaust on African consumer behavioural patterns. While this private gain further advances their profits, they inevitably contribute to Africa's development agenda through their investments in skills development programs for youth (in big data, cloud

computing and analytics) and in their application of big data to solve problems in sectors such as agriculture, energy and healthcare[9].

Another important group of stakeholders we note from our literature review are the international organizations and development partners. These tend to work in tandem with the government, civil society and private sector to fund and pilot several big data initiatives. Such important collaborations include that of the U.S. Agency for International Development (USAID) with various Ministries of Health and statistical bodies across Africa to collect health data and demographics through the Demographic Health Survey[10]. This has greatly improved the standards and methods for measuring nutrition, child and maternal health as well as various diseases. For instance, The UN hosts and supports three different big data initiatives – UN Global Pulse Lab, UN Big data lab, GPSDD while the CGIAR has established a Platform for Big Data in Agriculture.

Academic institutions on the other hand are working hard to up-skill new talent with the skills to support the adoptions of big data initiatives by offering degrees in Data Science and Artificial Intelligence. In several cases they are partnering with big tech companies originating from Silicon Valley such as Oracle, IBM, Microsoft, Google, Facebook and Amazon to offer capacity building by sponsoring meet-ups, workshops, trainings and in some cases, even university courses.

Academic Institutions such as Moi University and Jomo Kenyatta University of Agriculture and Technology in Kenya are also at the forefront in improving how governments adopt big data strategies by designing data collection devices that improve the time, accuracy and storage of data such as the ones used for the 2019 census in Kenya. Establishment of research facilities such as IBM Research Africa in partnership with local universities is also accelerating application of big data adoption in development projects across Africa[11].

9 Jilani, George. 2019. "Big Data: The Here and Now of African Healthcare." Diplomatic Courier. https://www.diplomaticourier.com/posts/big-data-the-here-and-now-of-african-healthcare.

10 Demographic Health Survey https://www.measureevaluation.org/resources/publications/ms-15-106/at_download/document

11 Jao, Nicole. 2017. "Big Data to tackle grand challenges: A look at IBM Research Africa's projects." ITU News. https://news.itu.int/big-data-to-tackle-grand-challenges-a-look-at-ibm-research-africas-projects/

Launch of the Africa Regional data Cube by African governments and the Global Partnership for Sustainable Development Data (GPSDD) in 2018.

BIG DATA INFRASTRUCTURE AND CAPACITY

Big Data for Development projects differ from standard ICT for Development projects because they require specialized skills and technologies. For instance, it is easier to deploy mobile-based development projects due to the fact that mobile phones are now ubiquitous and widely accessible to the general population without the need for specialized training.

WHO estimates that a big data project *"inherently demands more technical skills, specialized equipment, interoperability standards, coherent data collection and analysis systems and regulatory oversight"*[12]

As more and more sectors adopt digital tools, they are generating more data that domain experts can tap into and apply to their daily work. In fact, the biggest impact of a big data deployment is felt within sector verticals. However, these experts tend not to be trained in data analytics and therefore need to rely on working with data scientists to process and analyze the large volumes of data for decision making. This, therefore, creates an urgent demand for data scientists to support domain work with big data skills.

There is very limited technical capacity to manipulate big data using the specialized tools, algorithms and infrastructure across all African countries, just as it is elsewhere in the world. Where it exists, it is mostly to be found in the private sector and they may not be working deliberately to overcome existing development challenges. Several universities across the continent are adapting to this reality and creating advanced degrees that aim to nurture new data professionals. However, in several of these cases the most visible trainings seems to be conducted by non-academic institutions, and often in partnership with a private sector entity. For instance,

12 Wyber et.al 2015. "Big data in global health: improving health in low- and middle-income countries." Bulletin of the World Health Organization. https://www.who.int/bulletin/volumes/93/3/14-139022/en/

the Dara Big Data Africa summer school is a non-formal training on fundamental data science tools and techniques that is run annually by partners of the Square Kilometre Array program.

This comes as no surprise given the lengthy period of time that academic institutions require to develop academically rigorous curricula. Industry is moving at a fast pace and the demand for data proficient talent increases daily. Private companies, moreover, have gained experience from processing massive amounts of data and are therefore willing to support trainings to meet their industry gaps.

There is a commensurate need for rapid upskilling of African data professionals working towards meeting development goals in the continent over the next few years. This is necessary as the design of scalable solutions requires the input of people who understand local contexts and nuances of the populations they intend to impact. The development sector needs to adopt the agility of the private sector entities to ensure that they are adequately staffed with appropriately skilled professionals supported to drive the data revolution. This way, they can overcome the current scarcity of big data professionals.

There is a need for clean, and accurately labelled datasets to train models for big data and Artificial Intelligence applications. However, this data is very scarce and it is time consuming to re-purpose data collected for other reasons towards development focused projects. That said, the labelling of data to be used in big data applications often does not require specialized skills. In fact, we see western companies turning towards mechanical turks/ crowd-sourced initiatives to support these activities. These workers are often from developing nations where labour is not expensive. Samasource is an example of a company based in the US that recruits African workers to train data to be used in AI solutions. There is an opportunity for similar initiatives to hire local workers to support continental big data efforts. This is particularly advantageous as they are able to understand language contexts and label data than can be used in big data applications in local languages.

It is important even as we recommend the need for upskilling and developing more data science talent to approach big data applications in a multi-faceted by connecting and training these individuals by experts from other sectors such as social sciences, who can contextualize the knowledge and help them draw appropriate insights and conclusions from big data analysis.

Africa is still behind the curve and riddled with poor infrastructure and undeveloped hardware necessary to collect, mobilize, and share data is a crucial and neglected requirement paralyzing

the efforts of developing economies. There is a broad range of affirmative actions needed to build up the capacities of Least Developed Countries so that they can equally share and engage in data collection and sharing, rather than relying on the information altruism of stronger countries.

On a continent where electricity penetration is still not at 100%, basic infrastructure requirements to manage data centres and other big data computing facilities that require constant cooling may not be readily available. This limitation makes big data processing costly due to required access to the internet and servers that can manage the workload even as the unit costs have come down from previous years. With a limited number of data centres, there is latency and projects might take too long to execute or become too costly to run without significant upfront investments. Institutions such as Amazon, Microsoft and Icolo have established local data centres capable of storing and processing massive amounts of data with aim of reducing cost of managing big data applications. It is, therefore, important for governments and public private partnerships to be established to fund and further enhance underlying infrastructures for big data projects.

H.E. Paul Kagame, President of the Republic of Rwanda,is joined by other dignitaries at the opening of the Africa Drone Forum held in Kigali in February 2020. Rwanda has taken steps to create an enabling environment for new technologies such as drones and big data.

ENABLING ENVIRONMENT FOR BIG DATA IN AFRICA

With the eagerness to use big data as a resource to accelerate development outcomes, it is important to establish appropriate mechanisms to govern the usage of big data. Just like other new tools and technologies, big data use needs to be protected and streamlined by policies and laws governing its creation, consumption, distribution, implementation and validity.

A good enabling environment facilitates improved data collection methodologies, including the use of citizen-generated data and alternative datasets from the private sector, to supplement official statistics in cases where they are not being collected or where weak statistical systems exist.

The analysis of anonymized data has been in practice for a long time, but with the ability to offer personalized services to millions of people concurrently due to the computational prowess of big data technologies, individually identifying data is now commonly used in transactional systems as big data. Private companies typically harvest and own most of this valuable big data, and in some cases, sell them to third parties, who may not be domiciled on the continent, commoditizing individuals' information and likely infringing on people's privacy if no consent is sought from the individuals who the data is about.

Therefore, due to the increasing use of personal data in big data use cases, there is a need to ensure individuals' privacies are maintained and prevent data from misuse, threats to personal safety or promoting discrimination against groups of people. This is even more imperative when handling personal data of people from vulnerable and marginalized who may not be empowered to demand for their right to privacy.

Across Sub-Saharan Africa, most countries are working towards implementing various legislations and policies to regulate big data use. According to the Privacy International, 24 African

countries have adopted data protection laws and regulations to preserve the right to privacy for individuals[13]. Several of these laws, however, are fashioned against the European Union GDPR, which might be expensive or difficult to enforce. It is therefore important for African countries to review their data protection laws and support them with regulations that are fit for the contexts in which they are being enforced.

In the same vein, we need to rethink traditional consent. In the wake of massive data collection exercises such as drone mapping and satellite imaging, private companies and governments are now necessitated to generally enforce an ethical code of conduct when using big data pertaining to individuals' personal information. This code of conduct ideally should complement informed consent seeking by requiring companies to emphasize more on data governance after collection in terms of appropriate use and minimizing risk of breaches. This requires a constant commitment to anonymizing data. Further this data governance can be concretely implemented through openly publishing the adopted codes of conduct and data governance policies to promote trust from subjects of these big datasets.

Additionally, given limitations to capacity and infrastructure in developing countries in the Sub-Saharan region and more generally the global south, often times, external parties and multi-national organizations are largely involved in the collection and analysis of big data e.g. mapping projects. This raises new concerns over threats to national security and state sovereignty, especially if this data is not processed by representatives of the people it concerns. It is therefore important for African governments and development partners, to make a concerted effort to build clear data sharing strategies to address these concerns. One of the ways that this is promoted is the requirement that data pertaining to citizens are domiciled in countries where they are produced.

Proactive disclosure of data in open formats from governments go a long way in addressing these concerns, as there are clear guidelines as to how the data is produced and limits exploitation of individuals as they can keep parties accountable and increase trust when they know that they can access the data themselves and make decisions about their use. Open Data further benefits development agendas as it promotes data sharing between state agencies and ministries, and offers accountability to citizens on development projects.

There needs to be deliberate "gendering" of big data to ensure that women and marginalized groups are not left out – ensuring that big data projects do not perpetuate existing inequality

13 Privacy International. 2020. *2020 is a crucial year to fight for data protection in Africa.* https://privacyinternational.org/long-read/3390/2020-crucial-year-fight-data-protection-africa.

and bias in service to these groups. Big data projects should therefore support representation of all citizens, as the data used is collected at a granular level. This requires the government to invest into new data collection tools and methodologies that take into account and place in focus these minority and disadvantaged communities. This also requires a different approach to funding big data for development use cases.

Currently, big data projects tend to be funded pilots that rarely progress beyond the pilots as ROI, especially for development projects may not be immediately seen as it is for other public services. However, for governments to fully realize the potential benefits for big data projects in development, they need to create funding mechanisms – be it within government agencies such as the statistical bureaus, or to support public-private partnerships - that are earmarked for the medium-term to long-term and can support continued realization of impact even after donor funding is completed.

BIG DATA FOR DEVELOPMENT USE CASES

Although the use of big data in low- and middle-income countries has not taken off as it has in more developed countries due to the infrastructural and capacity challenges identified, it still poses an opportunity to transform societies by improving on traditional approaches to development. In this section we highlight emerging impacts of big data for development initiatives and how they are already shaping the future of economic sectors and governance in Africa.

Health Care

Many African countries aspire to quality healthcare and universal coverage for their population. By harnessing and analyzing big data, more targeted and efficient interventions can be made to populations at risk. We are seeing emerging use cases where big data is applied to design tailored care regimens for individuals based on their genetics and medical history. With the use of mobile phones, community health workers can also now offer targeted services remotely to patients.

For instance, the IBM Lab in South Africa has been developing an algorithm that identifies genes that cause cancer cells to spread in African patients. The lab has also modelled data from the World Health Organization (WHO) to better predict incidences of the tropical disease, malaria, and help government control spread of the disease based on among others, climate data.

The GPSDD has also been working with partners to develop standards for interoperable healthcare data that accelerates access to healthcare as a patient's medical history can be accessed by physicians and government health workers without the patient being treated as a new patient each time they seek treatment.

Another example of big data use in healthcare is in the re-purposing of mobile data to study the spread of epidemics. In 2014, Orange (the telecommunications company) in Côte d'Ivoire

released anonymized mobile phone data which was used by partners to develop a model for migration patterns and therefore, possibility of disease outbreaks and where they were likely to spread to[14]. A similar approach was used in West Africa to control the spread of the Ebola outbreak. Using mobile-phone-based surveillance, health workers were able to report progression of the disease to institutions such as the CDC and contain the outbreak[15] using readily available mobile technology.

Big Data and digital data collection systems are also expected to support governments with poor CRVS systems to easily collect statistics on births, immunization records, nutrition records and deaths of young children in order to ensure appropriate and timely interventions that prevent death from communicable diseases[16]. At a population level, traditional health data includes information from vital statistics registries and hospital admission statistics.

These digital records support analysis of data for important insights into threats to the general health of a section of the population. This can then inform new health strategies and influence better decision making and resource allocation for governments and private stakeholders.

Agriculture

The use of big data in agriculture is seen to overcome some of the challenges currently faced by the sector in Africa. Majority of farming on the continent is done by smallholder farmers (farming on less than an acre of land) who are currently not able to mechanize and optimize efficiency when planting and harvesting or adequately adapt climate smart agricultural practices to overcome environmental shifts precipitated by climate change.

The concept of precision farming is set to revolutionize agricultural productivity by allowing farmers and policy makers to track and predict accurately various agricultural activities such as changes in planting and harvesting cycles, seed varieties, soil quality, pest management. Farmers can then maximize yields by providing crops with their exact requirements and adopting the best agronomic practices to address insights coming from the data analyzed.

An example of precision agriculture in practice is the FarmBeats product by Microsoft, a tool that tracks soil quality thus helps farmers to know the correct time to apply water, pesticides

14 Kirkpatrick, Robert and Vacarelu,Felicia. n.d. *A Decade of Leveraging Big Data for Sustainable Development.* https://www.un.org/en/un-chronicle/decade-leveraging-big-data-sustainable-development.

15 Etang, Alvin and Himelein, Kristen. 2019. *Monitoring the Ebola Crisis Using Mobile Phone Surveys.* https://link.springer.com/chapter/10.1007/978-3-030-25120-8_2.

16 Jilani, George. 2019. *"Big Data: The Here and Now of African Healthcare."* Diplomatic Courier. https://www.diplomaticourier.com/posts/big-data-the-here-and-now-of-african-healthcare.

and other farm inputs. Aerobotics, a South African agri-tech start-up has also developed a platform that uses data from drone images and machine learning to help farmers identify pest and disease issues in crops[17].

Farm Drive uses data from mobile phones and satellite images to build a history and an identity for small-scale farmers, which makes it easier for either the government or the private sector to extend credit facilities to them. Their data can also be used to inform government agriculture policies as they are able to track which farmers are growing which crops and the quality of yield.

Energy

In order to support the African development agenda of inclusive growth, the African Development Bank is scaling up investment and implementation on five priority areas, referred to as the High 5s, with the Energy sector being one of them – lighting up and powering Africa. This means reducing carbon emissions by way of waste and instead investing in the efficient production of clean renewable energy as a means of sustainable development. Big data therefore, provides a useful tool and resource to achieve this objective as illustrated by the following use cases.

M-kopa, one of the largest providers of solar home systems use IOT and mobile based solutions to monitor usage of their kits and enable pay-as-you-go instalments from their users. Similarly, Blodgett et al. discuss the development of microgrids in East Africa that use the IoT to monitor performance of units and accurate metering in real-time. This results in targeted supply and establishment of efficient maintenance for the systems even in the most rural communities[18].

Nigeria's Kano Electricity Distribution Company (KEDCO) used tools such as Open Street Maps (OSM) to map their grid lines, and equipment such as transformers and substations to obtain metrics of the length of the grid lines and model the electricity loads and demands in certain places. This is in turn supported their overall maintenance planning[19]. Similarly, in Kenya, the main electricity distributor, Kenya Power, deployed an automated system that would not only consolidate customer data collection from 10 different sources, but also mine and analyze cus-

17 Forbes Africa. 2018. *"Big Data And Smart Farmers For Africa's Agricultural Transformation."* https://www.forbesafrica.com/focus/2018/10/30/big-data-and-smart-farmers-for-africas-agricultural-transformation/

18 Blodgett, C., Moder, E., Kickham, L., Leaf, H., 2016. *Powering Productivity. Early Insights into Mini Grid Operations in Rural Kenya*

19 Onyeji-Nwogu, Ijeoma, Morgan Bazilian, and Todd Moss. 2020. *"Big data and the electricity sector in African countries."* Science Direct https://doi.org/10.1016/j.wdp.2020.100170.

tomer data. The analytics solution gives Kenya Power the ability to perform complex queries on data to give better insights on the varying needs of customers across different regions.

Financial Services

Improvement to Africa's financial inclusion can directly be attributed to the mobile revolution and to applications of big data technologies. According to the World Bank, between 2011 and 2018 the penetration of financial products and services among adults in sub-Saharan Africa rose from 24% to 43%[20].

This growth in financial inclusion is further attributed to more accessible product offerings from mobile network operators and alternative financial institutions that have enjoyed easier regulatory frameworks than those set for existing commercial banks in the continent. These have relied on studying thousands of data points on customer preferences and usage of financial products to either extend loans or savings products directly from mobile phones or from bank network agents at the last mile. One of the most recent big data innovations that has shown early success is M-Shwari and Fuliza[21], credit products from Safaricom and NCBA in Kenya. These two products analyze big data on M-Pesa mobile money transactions to create credit and savings products for subscribers. Other institutions such as Farm Drive are also using this mobile transaction data to create credit profiles for small-holder farmers for input and harvest loans, a section of the economy that is typically not able to access formal banking facilities.

Governance

Good governance practices enable governments to remain transparent, accountable and deliver services efficiently to citizens. We are seeing more and more African governments not only investing in e-governance solutions but also harnessing digital records from these systems Big Data analytics to offer improved services to their citizenry. Luvembe et al. posit that E-governance reduces frequencies of extortion cases, increases openness, clarity, stirs economic growth and decreases time wastage[22]. For instance, they illustrate that Kenya has re-purposed her Integrated Taxation Management Systems (TIMS) and Financial Management

20 World Bank. n.d. *"Digital Access: The Future of Financial Inclusion in Africa ."* http://documents.worldbank.org/curated/en/719111532533639732/pdf/128850-WP-AFR-Digital-Access-The-Future-of-Financial-Inclusion-in-Africa-PUBLIC.pdf.

21 Kanyi, Mark, and Kenneth Wanjau. 2019. *"How to turn Big Data into a solid source of innovation and solutions."* Standard Digital. https://www.standardmedia.co.ke/business/article/2001324420/how-to-turn-big-data-into-a-solid-source-of-innovation-and-solutions.

22 Elezaj, O., Tole, D. and Baci, N. (2018) Big Data in E-Government Environments: Albania as a Case Study. Academic Journal of Interdisciplinary Studies, 7, 117-124.

System (IFMIS) to analyze revenue efficiency and mitigate gaps in revenue collection in Kenya's county governments.

Digital Identity

Increasingly, more and more African governments are investing in Digital Identification programs aimed at unifying records held by government on their citizens under one identifier in order to improve service delivery to all citizens[23]. Countries such as Kenya and Cape Verde aim to deliver more services to citizens via means of ICT platforms. These digital identities are connected to an individual's respective biometric information allowing governments to monitor health and social data.

Urban Planning and Development

As urban ecosystems become digitized, big data use cases are emerging and achieving early traction in society. For instance, in the face of recurrent water shortages, a Kenyan GIS company, Upande, has deployed smart meters and sensors to support water companies and institutions utilize big data to monitor leakages and fraud in their water supply systems. In Ghana, satellite data made available via the African Regional Data Cube is used to monitor illegal mining and other encroachments on public land[24].

IBM has been working closely with city officials in Kenya and Cape Town in a bid to support their resource planning and efficient delivery of services. For instance, in Nairobi, IBM connected garbage collection trucks with sensors that were able to collect data not only of waste management but also of road quality. Similarly, in Cape Town, the Research team deployed a dashboard that draws together weather data and open data from the city's portal to monitor and assess the risk and severity of wildfires and manage emergency response to these incidents.

Trade and Regional Integration

The African Union, through its Agenda 2063 aspires to greater regional integration to ease movement of people and regional trade, thus increasing economic productivity. One of the

23 Luvembe, A.M. and Mutai, H. (2019) Big Data Framework for Kenya's County Governments. Journal of Computer and Communications, 7, 1-9.

24 AWS Public Sector Blog Team. 2018. *"Achieving Sustainable Development in Africa with Open Data ."* AWS Public Sector Blog. https://aws.amazon.com/blogs/publicsector/achieving-sustainable-development-in-africa-with-open-data/

goals was to establish the Continental Free Trade Area by 2017, which is expected to be in effect from 2020[25]. This is important as trade is seen as a key driver for sustainable development.

The Economic Commission for Africa (ECA), the African Union and the African Development Bank have jointly developed a Regional Integration Index[26] which aims to monitor the impact of regional integration. These institutions foresee big data such as airline data being analyzed for flight patterns between African airports and trade tariff data to be employed to monitor new indicators in the regional integration index. This provides an opportunity for instance, for entrepreneurs to assess the viability of expanding to new markets.

As part of regional trade expansion, big data can thus be leveraged to foster entrepreneurship, support sustainable trade practices and transform economies. It was speculated that retail businesses such as Nakumatt, a Kenyan supermarket and one of the largest contributors to the East African economy, failed to utilize the data it has gathered from its customers as one of the main reasons for its financial problems and ultimate failure. However, on the other hand, IBM has been working in collaboration with Kenyan government, by using machine-learning technologies to simplify the process of starting a business by reducing the number of interactions with the government from 11 to 3 steps and therefore improved the ease of doing business in the country.

25 *"African Continental Free Trade Area (AfCFTA) Legal Texts and Policy Documents."* Tralac. https://www.tralac.org/resources/our-resources/6730-continental-free-trade-area-cfta.html

26 Lopes, Carlos. 2015. *"The Role of Big Data in Africa's Regional Integration."* Africa Regional Integration Index. https://www.integrate-africa.org/news-knowledge/news/the-role-of-big-data-in-africas-regional-integration-60/

PARTNERSHIPS FOR BIG DATA FOR DEVELOPMENT

Successful big data initiatives require a combination of deep domain expertise, a robust regulatory framework, data science and big data manipulation skills and, active evaluation of the initiative impacts on society.

This creates a large scope of work for initiatives that intend to scale and have massive impact. In most cases, it is not feasible for just one institution to have all the expertise to roll out such a project. Therefore, as has been seen in the ecosystem mapping, there is a need to develop collaborations and cooperation of different stakeholders and tap into their collective expertise.

This requires building systems and communities of practice around data of thematic interest. The ADRR report refers to these collectives as data communities. Data communities bring together different stakeholders to work together to solve a common thematic issue. One of the proposed ways such a multi-stakeholder grouping works together efficiently is through the formation of data collaboratives – that is mutually beneficial groups of people coming together to exchange their data to create public value[27]. These partnerships and data exchanges can ensure timely data is made available for use and provide the requisite skills and normative frameworks to ensure the success of such initiatives.

Data collaboratives need to define how their work and cooperation is governed, which requires a clear understanding of the motivations, incentives and benefits of each stakeholder in the collaborative. Data communities require openness and transparency, in order to ensure that the mutual benefits of the collaborative are gained in an optimum manner.

THE NEED FOR A NETWORK

A network like BD4D is well suited to support emerging collaboratives by leveraging on its convening power, skills from each of the networks members and collating relevant research and

27 Data Collaboratives https://datacollaboratives.org/

resources to support activities of the various partners and emerging data communities. We envision the Africa Hub of the network offering support to existing and potential stakeholders aiming to use big data in the context of development as follows:

» Advocating for the development of favourable legislation and an enabling environment for big data for development initiatives to thrive by providing necessary evidence of big data innovation, data justice issues and their impact to local communities.

» Fostering innovation using big data that accelerates achievement of national development agenda by bringing together unusual partners to brainstorm and design solutions to development needs that account for local contexts, nuances and societal norms.

» Amplifying efforts of capacity building institutions and implementing partners of the various big data initiatives by connecting resources to those who require them.

» Supporting peer-learning from south-to-south collaborations and emerging big data collaboratives by documenting and introspecting on the various success, failures and limitations to these partnerships forming.

Ultimately, a big data for development network in Africa can achieve the above objectives by developing relevant research programs, pooling funds, technology and other required resources for network members to access and, amplifying publications and outputs emerging from various projects implemented by network members.

REFERENCES

n.d. “African Continental Free Trade Area (AfCFTA) Legal Texts and Policy Documents.” Tralac. https://www.tralac.org/resources/our-resources/6730-continental-free-trade-area-cfta.html.

2017. “African Governments Lead the Way on Data Revolution.” Global Partnership for Sustainable Development Data. http://www.data4sdgs.org/news/african-governments-lead-way-data-revolution.

Anderson, Mark. 2017. The hunt begins for Africa’s big data - The Africa Report. https://www.theafricareport.com/791/the-hunt-begins-for-africas-big-data/.

AWS Public Sector Blog Team. 2018. “Achieving Sustainable Development in Africa with Open Data .” AWS Public Sector Blog. https://aws.amazon.com/blogs/publicsector/achieving-sustainable-development-in-africa-with-open-data/.

Blodgett, Courtney, Emily Moder, Lauren Kickham, and Harrison Leaf. w016. “POWERING PRODUCTIVITY: Early Insights into Mini Grid Operations in Rural Kenya.” https://sun-connect-news.org/fileadmin/DATEIEN/Dateien/New/Kenya-Mini-Grids-White-Paper-VI2.pdf.

Dorian, Abbey. 2019. “Big Data And The Sustainable Development Goals.” The Organization for World Peace. https://theowp.org/big-data-and-the-sustainable-development-goals/.

ECA, UNDP, Open Data for Development Network, World Wide Web Foundation. 2016. “The Africa Data Revolution Report.” https://www.uneca.org/sites/default/files/uploaded-documents/ACS/africa-data-revolution-report-2016.pdf.

Etang, Alvin and Himelein, Kristen. 2019. Monitoring the Ebola Crisis Using Mobile Phone Surveys. https://link.springer.com/chapter/10.1007/978-3-030-25120-8_2.

Forbes Africa. 2018. “Big Data And Smart Farmers For Africa’s Agricultural Transformation.” https://www.forbesafrica.com/focus/2018/10/30/big-data-and-smart-farmers-for-africas-agricultural-transformation/.

GSMA Intelligence. 2020. “The Mobile Economy Sub-Saharan Africa 2019 .” https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/03/GSMA_MobileEconomy2020_SSA_Eng.pdf.

Jao, Nicole. 2017. “Big Data to tackle grand challenges: A look at IBM Research Africa’s projects.” ITU News. https://news.itu.int/big-data-to-tackle-grand-challenges-a-look-at-ibm-research-africas-projects/.

Jilani, George. 2019. “Big Data: The Here and Now of African Healthcare.” Diplomatic Courier. https://www.diplomaticourier.com/posts/big-data-the-here-and-now-of-african-healthcare.

Kanyi, Mark, and Kenneth Wanjau. 2019. “How to turn Big Data into a solid source of innovation and solutions.” Standard Digital. https://www.standardmedia.co.ke/business/article/2001324420/how-to-turn-big-data-into-a-solid-source-of-innovation-and-solutions.

Kirkpatrick, Robert and Vacarelu,Felicia. n.d. A Decade of Leveraging Big Data for Sustainable Development. https://www.un.org/en/un-chronicle/decade-leveraging-big-data-sustainable-development.

Lopes, Carlos. 2015. “The Role of Big Data in Africa’s Regional Integration.” Africa Regional Integration Index. https://www.integrate-africa.org/news-knowledge/news/the-role-of-big-data-in-africas-regional-integration-60/.

Onyeji-Nwogu, Ijeoma, Morgan Bazilian, and Todd Moss. 2020. “Big data and the electricity sector in African countries.” Science Direct https://doi.org/10.1016/j.wdp.2020.100170.

Privacy International. 2020. 2020 is a crucial year to fight for data protection in Africa. https://privacyinternational.org/long-read/3390/2020-crucial-year-fight-data-protection-africa.

Rao, Naveen. 2019. “Big Data Can Improve the Health of the World’s Most Vulnerable: Mothers and Children.” Time. https://time.com/5708751/big-data-maternal-health/.

n.d. “The State of Open Data: Histories and Horizons.” Open Data for Development. https://www.stateofopendata.od4d.net/.

van Rijmenam, Mark. 2013. “A Short History Of Big Data .” Datafloq. https://datafloq.com/read/

big-data-history/239.

Ward, Jonathan Stuart, and Adam Barker. 2013. “Undefined By Data: A Survey of Big Data Definitions.” Arxiv.org.

World Bank. n.d. “Digital Access: The Future of Financial Inclusion in Africa .” http://documents.worldbank.org/curated/en/719111532533639732/pdf/128850-WP-AFR-Digital-Access-The-Future-of-Financial-Inclusion-in-Africa-PUBLIC.pdf.

Wyber, Rosemary, Samuel Vaillancourt, William Perry, Priya Mannava, Temitope Folaranmi, and Leo Anthony Celi. 2015. “Big data in global health: improving health in low- and middle-income countries.” Bulletin of the World Health Organization , January.

www.developlocal.org