The Africa Data Revolution Report 2016

Highlighting developments in African data ecosystems

ECA

UNDP
Empowered lives.
Resilient nations.

OpenData
for development network

WORLD WIDE WEB
FOUNDATION

The Africa Data Revolution Report 2016

Highlighting developments in African data ecosystems

ECA

UNDP
Empowered lives.
Resilient nations.

OpenData
for development network

WORLD WIDE WEB
FOUNDATION

Contents

Foreword

African countries pledged to achieve sustainable development and inclusive growth for all following the adoption of the 2030 Agenda for Sustainable Development and the Agenda 2063. The need for timely and quality data to inform the two agendas exerts pressure on the national statistical systems more than ever before. While much progress has been made in past decades to strengthen national statistical systems for evidence-based decision making, countries still face challenges in many areas. The data required for monitoring progress toward the ambitious and aspirational global and continental development agendas is unprecedented both in scope and granularity with which the data should be collected.

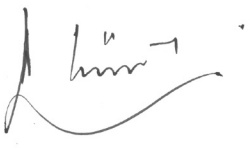
Technological advancement, growing recognition of individual rights for information and knowledge, the spread of open data initiatives, and increasing involvement of various non-traditional data stakeholders offer opportunities to meet data challenges for the implementation and monitoring of the development priorities at global, regional and national levels.

Indeed, the potential of data goes beyond reporting of development progress. Empirical evidence is mounting that data enables citizens to make more effective decisions in their daily lives, entrepreneurs to create new business opportunities, and institutions to make the governing process more efficient, responsive, inclusive and transparent. Nevertheless, data alone cannot unlock the above-mentioned opportunities if laws and policies allowing people free access, use and reuse of data are not in place. And such laws and policies mean little if technical capacity and infrastructure to make use of the data are not sufficient. Infrastructure would be useless if it is not adapted to local context, and if communities cannot access or afford it. Data communities would not fully grow and prosper if they are not empowered to voice their particular needs and capacitated to engage in partnerships and design their own innovative solutions.

To fully harness the data revolution, a holistic strategic approach is therefore required. Each and every component of the data ecosystem should be steered toward the shared goal of establishing a conducive environment for leveraging data and partnerships to respond to development priorities while leaving no one behind.

This first edition of the biennial Africa Data Revolution Report aims to address issues on the process of transforming data ecosystems, from the status quo to an aspirational state where data from both conventional and new sources are being harnessed to better inform decision-making and enable sustainable development with contributions from diverse data communities. The report is envisioned to assist countries on the continent in charting their way towards strengthened national data ecosystems by identifying common problems, sharing home-grown examples and lessons learned and coming up with actionable recommendations applicable in the regional, national and local contexts. Further, it reviews the current state of data ecosystems in Africa in terms of the diversity of data actors and their capacity needs, legislative and policy frameworks, technological infrastructure, tools and platforms, and the dynamic interactions between them.

We hope that this inaugural report will serve as a baseline from which countries will assess their progress toward harnessing the data revolution at the service of their development priorities, galvanize African-led regional efforts and feed into the discourse on data for sustainable development in global fora.



Oliver J.M. Chinganya

Director,
African Centre for Statistics
United Nations Economic Commission for Africa

Mansour Ndiaye

Team Leader,
Inclusive Growth and Sustainable Development
Regional Service Center for Africa
United Nations Development Programme

Fernando Perini

Coordinator,
Open Data for Development Network
International Development Research Centre

Anne Jellema

Chief Executive Officer,
World Wide Web Foundation

Acknowledgements

This report is the joint product of four collaborating institutions: the United Nations Economic Commission for Africa (ECA), the United Nations Development Programme (UNDP), the International Development Research Centre of Canada and the World Wide Web Foundation. These institutions were represented by a Steering Committee composed of Chukwudozie Ezigbalike and Inkyung Choi (ECA), Serge Kapto and Alessandra Casazza (UNDP), Fernando Perini (International Development Research Centre of Canada) and Nnenna Nwakanma (World Wide Web Foundation), together with the lead author of the report, Cosmas Ochieng of the African Centre for Technology Studies. The contribution made by UNDP was made possible thanks to the generous support of the Government of the Republic of Korea. The Open Data for Development is funded by the International Development Research Centre, the Government of Canada and the World Bank, all of which were indispensable for the production of this report.

Under the supervision of the Steering Committee, the report was put together by a team of country researchers headed by the lead author and comprising Jean-Paul van Belle (University of Cape Town, South Africa); Canisius Nzayisenga (Independent Consultant, Rwanda); Olusegun Akanbi (Independent Consultant, Nigeria); Solomon Selassie (Independent Consultant, Ethiopia); Kenneth Mdadila (United Republic of Tanzania); Paul Mungai (University of Cape Town, South Africa); Godwin Hlatshwayo, (Independent Consultant, Swaziland); Rakotomanana Eric (Independent Consultant, Madagascar); Suwadu Sakho-Jimbira (Initiative Prospective Agricole et Rurale, Senegal); and Charles Doukouré Fe (Independent Consultant, Côte d'Ivoire). The country researchers produced 10 background papers for this report. They received support from national statistical offices and UNDP country offices including those in Côte d'Ivoire, Ethiopia, Madagascar, Nigeria, Rwanda, South Africa and the United Republic of Tanzania. The country research in Nigeria was supplemented by a report on "Unpacking the data revolution in Nigeria" prepared by the Federal Public Administration Reform Programme, with permission from the United Kingdom Department for International Development. Country research in Senegal and Swaziland was reused from the UNDP Ecosystems Mapping Project. Solicited contributions were also received from Davis Adieno (Civicus, the World Alliance for Citizen Participation), Danielle Doughman and Donatien Beguy (African Population and Health Research Center), Paula Hidalgo–Sanchis (United Nations Global Pulse, Uganda) and Nnenna Nwakanma (World Wide Web Foundation).

This report benefited from an earlier consultative workshop in Nairobi, Kenya in August, 2016. Peter da Costa (Kenya), Inkyung Choi (ECA), Serge Kapto and Bouazza Bouchkhar (UNDP), Yosr Abdel Faatah Ahmed Ibrahim (CAPMAS, Egypt) and Steve Kenei (Development Initiatives) provided comments on an early draft of the report.

Acronyms

AfDB	African Development Bank
ECA	United Nations Economic Commission for Africa
ICT	Information and communication technologies
STEM	Science, technology, engineering and mathematics
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization

Executive Summary

The data revolution can be harnessed to catalyse positive social, economic and environmental transformation in Africa. Considerable innovation and experimentation is currently under way within multiple data communities and ecosystems in many African countries. However, for the most part, these are small-scale, pilot, isolated or ad hoc initiatives. If Africa is to benefit from the full transformative potential of the data revolution, more systematic, large-scale, integrated and sustainable efforts are going to be needed.

African governments and other stakeholders recognize the importance of the data revolution for the continent's development. At the continental level, this recognition is embodied in the Africa Data Consensus[1] and other statistical and development initiatives, including the African Charter on Statistics and the African Union's Agenda 2063. At the national level, this can be seen in long-term national development plans and numerous legal, legislative and policy reforms aimed at improving the quality, timeliness, relevance, availability and accessibility of data. In many countries, these national and continental development goals are in the process of being aligned with the global development priorities to which most African countries have subscribed, including most notably the United Nations Sustainable Development Goals. This alignment entails efforts to harmonize, coordinate or integrate data requirements for the Sustainable Development Goals and national development priorities.

The building blocks for a data ecosystem which is capable of harnessing the data revolution for accelerated sustainable development already exist in Africa. These include multiple dynamic and growing data communities which range from official statistics and private-sector, civil society and citizen-based data groups to scientific, open and big data communities. They also include legal, legislative and policy frameworks aimed at creating enabling environments and governance frameworks for harnessing the data revolution at the national and regional levels, such as the African Charter on Statistics and the Strategy for the Harmonization of Statistics in Africa. A data ecosystem which is capable of harnessing the full transformative potential of the data revolution also requires appropriate technological infrastructure and architecture and human and financial resources. The rise of information, communication and geospatial technologies, among others, together with demographic changes, Africa's recent economic growth rates and increasing democratization provide foundational structures for data ecosystems capable of harnessing the data revolution.

In spite of this, significant deficits, gaps and weaknesses remain within the current data ecosystems in many African countries. If Africa is to harness the full transformative power of the data revolution, significant investment will need to made in human resources, technological capabilities, platforms and tools, and in establishing effective governance frameworks related to the production, processing, protection, ownership, quality, openness, timeliness, relevance, accessibility, harmonization, interoperability and use of different types of data, regardless of who produces or owns them.

This report examines the current state of the data ecosystem in Africa, its desired end state, and the gaps in between. It also provides recommendations on how to bridge these gaps. It maps the current data ecosystem in Africa in terms of purpose, actors, principles and protocols; legal, legislative and policy frameworks; technological infrastructure, tools and platforms; and the dynamic interactions between purposes, actors, frameworks, technologies and systems.

This report conceptualizes the "African data revolution" and identifies what is revolutionary about it – notably (a) the exponential increase in the volume, types and speed of data available in African countries; (b) the increased availability and use of new types of data as well as new uses of old or conventional types of data; (c) new principles of data governance, including the principles of inclusion and openness; and (d) new data-related rights and free-

1 United Nations, Economic Commission for Africa, 2015.

doms. The report argues that underlying the concept of the African data revolution is 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.

The question of what and who officially counts and is counted, by whom, how and for whom is more than simply a question of available or feasible statistical techniques and methodologies. At its core, this question embodies political economy and ideological considerations of what and who constitutes a national priority, and is worthy of counting, and by extension, planning public, social, development and other services for. Official statistics can be and have been used as tools for social, economic and political exclusion and segregation in Africa, especially under colonial and apartheid rule. However, statistics and data in general can also be used as tools for promoting social, economic and political inclusion and integration, for example by including marginalized communities, regions and peoples in national censuses, household surveys and administrative, civil and other vital data registration exercises. The report concludes that the African data revolution embraces a broader concept of who and what officially counts, is counted, by whom and for whom. This derives in part from the broader development goals for which African countries seek to harness the data revolution: the Sustainable Development Goals adopted by the United Nations and the African Union's Agenda 2063.

The report makes the following recommendations:

Proof of premise: What is the full range of social, economic, environmental and other impacts – positive and negative – of the data revolution? What might African countries do to maximize the positive and mitigate the potentially negative impacts of the data revolution?

The full range of socioeconomic and other impacts of the data revolution and the various types of data such as big data and open data remains largely unknown. Without this knowledge, what interventions governments and other stakeholders might undertake to maximize the benefits of the data revolution and minimize its negative potential remains limited. In order to accelerate the establishment of data ecosystems which are capable of harnessing the data revolution, it would be helpful if policymakers and other decision makers had a clear sense of the full range of impacts of different types of data, data platforms, technologies and tools. The scientific community and the international and continental development bodies such as ECA, the African Development Bank (AfDB) and UNDP can help in commissioning, financing or undertaking these scenario studies and impact studies. If data, including new and novel types of data, are to be used in decision-making, the case for many of the claims made on behalf of such data must be backed up by solid evidence. This is likely to enhance public investment in national data ecosystems, but also the use of data produced by such systems. Africa must create a virtuous cycle of good-quality data, good decision-making, good policy outcomes and sufficient investment in national data ecosystems if it is to harness the data revolution on a large scale and sustainably.

The importance of capacity needs assessments early in the cycles covered by the Sustainable Development Goals and Agenda 2063

It is almost taken for granted that every African country is ready for different types of data, technologies, tools and platforms. Yet there is widespread recognition that despite recent improvements, the national statistical system in many African countries is capacity-constrained. Different data communities might have arisen in different African countries, but the assumption that every African country is ready to harness different types of data ignores fundamental differences in political economies and institutional, technological, financial and human resource capabilities. Whether a country is in a position to effectively harness geospatial and big data, for example, depends on a whole range of factors that cannot be taken for granted. If African countries are going to establish data ecosystems which are capable of harnessing the data revolution, national assessments of readiness to handle different types and sources of data, technologies and platforms are going to be critical in helping them to identify gaps or areas where key interventions might be directed.

Legal, legislative and policy reforms to anticipate and allow for the harnessing of the data revolution for accelerated sustainable development in Africa

A number of reforms have been initiated in many African countries, but in many cases they are either incomplete or insufficient. The right legal, legislative and policy environments will create conditions that are conducive to harnessing the data revolution for sustainable development in Africa.

Significant investment in data technologies, platforms and tools and human and financial resources is going to be needed if Africa is to benefit from the full transformative potential of the data revolution

This includes general investment in information and communication technologies (ICT) and infrastructure, including most notably Internet, mobile and digital technologies. It also includes investment in data science, computing, STEM (science, technology, engineering and mathematics), statistics, social sciences and economics skills at a broader tertiary level and with specific reference to the national data ecosystem.

Greater collaboration and coordination between data communities can significantly reduce the costs of data collection for use in pursuing the Sustainable Development Goals, Agenda 2063 and long-term national development plans by helping fill gaps in official statistics and enhance data quality, timeliness, relevance, accessibility, dissemination and use.

Data collection is costly. National statistical offices are unlikely to have the necessary and sufficient technological, financial and human resource capacities to collect, process and disseminate the data required for the pursuit of the Sustainable Development Goals, Agenda 2063 and national development plans. However, significant capabilities and resources are available within other data communities in Africa. These include private-sector, civil society and citizen-based data communities and scientific, open and big data communities. Greater collaboration, coordination, harmonization and integration of data initiatives among these communities have the potential to significantly reduce the costs of data collection and to fill key gaps in national official statistics. They also have the potential to enhance data accessibility, dissemination and use. Public-private and other partnerships, and other business models for facilitating collaboration in data collection, processing, analysis, dissemination and storage, need to be explored.

The importance of strengthening investment in administrative data collection and use

Administrative data, including civil registration data, are critical in national development and planning. Civil registration is the means by which countries keep continuous and complete records of births and deaths. The issue and possession of birth and death certificates dictate several civic, professional, business and personal activities and transactions. In many countries, it is through civil registration documents that individuals are admitted into schools and hospitals, gain nationality and formal employment, vote or present themselves for electoral offices, buy and transfer properties and access financial and legal services. To be excluded from civil registration, then, is almost tantamount to being excluded from public services. In spite of this, 46 African countries do not have complete civil registration systems to register births. If the continent is going to harness the data revolution to accelerate sustainable development, a majority of African countries are going to need to strengthen their capacities to collect and use high-quality administrative data.

National statistical offices are and will remain the cornerstone of national data ecosystems

In spite of - in fact because of - the rise of multiple data communities, the national statistical system will continue to play a central role in national data ecosystems in many African countries. It can be a key facilitator and leader in fostering collaboration, harmonization and coordination within national

data ecosystems. The other data communities will continue to complement rather than supplant the national statistical system. How the official statistics community relates to the other data communities will to a large extent determine the degree to which African countries harness the data revolution.

Functional autonomy for the national statistical system

Owing to its continuing centrality within the national data ecosystem, the national statistical system requires functional autonomy, or a lack of politicization of its work, coupled with sufficient financial and human resources. It cannot be expected to collaborate with the other data communities or to facilitate collaboration, harmonization and coordination if its technical and technological capabilities are substantially weaker than those of the other data communities, or if it does not have sufficient financial resources to undertake its core mandate, which would form the basis of any collaboration, harmonization and coordination.

Non-State-based data communities have a responsibility to act proactively and responsibly to contribute to the creation of functional, vibrant, harmonized and integrated national data ecosystems

Legal, legislative and policy reforms coupled with advances in ICT have opened up and or are opening up the data ecosystem in many African countries to non-State-based data communities. The opportunity to seize this growing policy, legal and legislative space is increasingly available to non-State-based data actors and communities. To be sure, many constraints to participation by non-State actors in national data ecosystems remain. However, the policy, legislative and legal openings are there. Non-State-based data communities can seize these opportunities not only to hold governments accountable through the open data movement, but also to share data, improve data quality and use, and proactively seek collaboration with the official statistics community.

Learning from experimentation and innovation

Realizing an effective, responsible, coordinated, harmonized and integrated national data ecosystem will require significant experimentation and innovation, and efforts to learn from them. There is nothing automatic about the data revolution. It is the result of deliberate legal, policy and technological choices and actions. Successes or failures of experimentation and innovation will be constrained by these choices and structural and external factors, among others. The ability to identify the reasons behind successes and failures within different data communities, and to apply these lessons in future interventions, will determine the type of data communities and ecosystems that emerge in Africa in the future, and the degree to which the continent harnesses the full transformative power of the data revolution. The reality of experimentation and innovation will need to be appreciated especially by the public sector, where failure is not always tolerated. Innovative models for conducting experimentation and innovation such as incubator labs or pilot projects may need to be explored before large-scale projects are rolled out.

1. Introduction: The African data revolution and sustainable development

The data revolution can be harnessed to accelerate sustainable development in Africa. Data are available ever faster, in greater volume and scope. Globally, the quantity and quality of data on social, economic, political and ecological activities are expanding exponentially. This is driven by growth in the numbers and analytical power of computers, machines and mobile devices; the spread of the Internet; the falling cost of data storage; and the surge in remote sensing, satellite and other technologies that are creating new data and tools on an unprecedented scale. It is also driven by a surge in demand for data from the public and private sectors and civil society, and by greater experimentation, innovation and adaptation in data production, dissemination and governance across multiple data communities.

In its Data Manifesto,[2] the Royal Statistical Society in the United Kingdom argues that what steam was to the nineteenth century, and oil to the twentieth, data are to the twenty-first. According to the Society, data are the "driver of prosperity, the revolutionary resource that is transforming the nature of social and economic activity, the capability that differentiates successful from unsuccessful societies". A number of recent scholarly, policy and commercial reports reinforce this characterization. Goldman Sachs observes that while economic productivity doubled in 150-200 years during the industrial revolution, it "skyrocketed in the computer revolution, with a three-fold gain in the half-century starting in the 1960s". The firm believes that with the "fourth revolution" (i.e. the data revolution), the world may have entered a period of even more rapid productivity gains.[3] Reports by the World Economic Forum, the McKinsey Global Institute, the EMC Corporation[4] and a number of academics describe huge potential benefits of the data revolution for entire economies and sectors such as health, agriculture, industry, transport, retail and public services. There is growing consensus that while the data revolution does not constitute the first time that technology has promised to transform the world, this process might be the fastest.

Big data has been variously described as the oil of the digital economy, the next big thing in medical care and an integral tool for smart city planning and development.[5] The global open data community makes similarly strong arguments for the potential

2 Royal Statistical Society, 2014.
3 Goldman Sachs Asset Management, 2015.
4 EMC Corporation, 2014. See also Connelly and others, 2016; Einav and Levin, 2013; Einav and Levin, 2014.
5 de Montjoye, Y.-A., 2014.

applications of open data in enabling innovations that are catalysing social, economic and political transformations by changing sectors as diverse as health, education, urban development and fiscal governance. Openness is seen as foundational in harnessing the opportunities for data innovation that are unlocked when data cross geographical, territorial, sectoral, disciplinary, social and other boundaries.[6] Data generated by citizen groups have also been shown to hold great promise, while the power of high-quality official statistics and scientific data in helping inform policymaking has long been recognized.

African governments, civil societies, academics and private-sector organizations have recognized the potential applications of the data revolution to the continent's sustainable development efforts. Jakaya Kikwete, the former President of the United Republic of Tanzania, spoke for many African governments when he opened the first ever Open Data Conference in Africa in Dar es Salaam on 4 September 2015: "I totally subscribe to the idea that data is an important tool for development. No successful and meaningful development will take place without the use of data … We cannot honestly speak of developing Africa today without improving on ways we source, process and make use of data. More importantly, we must transform ways we manage data by increasingly making public data open and shared. In that way we will enable decision makers to make informed decisions. We will also empower the public to use data to hold the public officials accountable. In that way we will enhance good governance and take Africa to the next level."[7]

In its Data Revolution Policy and *SMART Rwanda Master Plan 2015-2020*, the Government of Rwanda makes the case that the explosion in the world of data has "enormous potential to spur innovation and creativity, increase the proliferation of highly skilled jobs, contribute to economic growth, enable better decision-making and create a more accountable, efficient, responsive and effective government". The master plan commits the Government to develop a policy on Open Data, on the basis that "easy access to data allows individuals and organizations to develop new insights and innovations that can improve the lives of others and help to improve the flow of information, spur development, promote transparency and build capacities". Similar arguments for the potential benefits of the data revolution for sustainable development in Africa have been made by various African governments in their national development strategies and by African scholars and development agencies in various books and papers, including Ben Kiregyera's *The Emerging Data Revolution in Africa*; the Mo Ibrahim Foundation's *Strength in Numbers: Africa's Data Revolution*; and Johan Fourie's *The data revolution in African economic history*.

1.1 Defining the data revolution in Africa

The United Nations Independent Expert Advisory Group on the Data Revolution for Sustainable Development defines the data revolution as an "explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies such as mobile phones and the 'Internet of things', and from other sources, such as qualitative data, citizen-generated data and perceptions data".[8] The Group considers that the data revolution includes a "growing demand for data from all parts of society". The data revolution for sustainable development, in the Group's view, involves "the integration of … 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 on individuals and groups, and minimising inequality in production, access to and use of data; … more empowered people, better policies, better decisions and greater participation and accountability, leading to better outcomes for people and the planet".

6 Third International Open Data Conference, 2015.
7 United Republic of Tanzania, 2015.
8 United Nations, 2014.

The Africa Data Consensus, developed at the High Level Conference on Data Revolution, held in Addis Ababa, Ethiopia, from 27 to 29 March 2015, builds on the Group's definition. It defines the African data revolution as: "A profound shift in the way that data is harnessed to impact on development decision-making, with a particular emphasis on building a culture of usage. The process of embracing a wide range of data communities and diverse range of data sources, tools, and innovative technologies, to provide disaggregated data for decision-making, service delivery and citizen engagement; and information for Africa to own its narrative". The Consensus views the data revolution as a "partnership of all data communities that upholds the principles of official statistics as well as openness across the data value chain, which creates a vibrant data ecosystem providing timely, user-driven and disaggregated data for public good and inclusive development".[9]

The Consensus is a strategy for nurturing the data revolution in Africa. It aims to create a new data landscape or ecosystem by opening up the field of data production and dissemination to State and non-State actors. It was developed in response to calls for a framework on the data revolution in Africa and its implications for the African Union's Agenda 2063 and the Sustainable Development Goals. The plan of action for the application of the Consensus is being spearheaded by ECA, the African Union Commission and AfDB, with support from UNDP and the United Nations Population Fund, and in collaboration with partner institutions from the public and private sectors as well as civil society organizations.

The core principles of the Consensus include:

Box 1. Principles of the Africa data revolution

Data must be disaggregated to the lowest levels of administration by gender, age, income, disability and other categories.
People must be counted to make them count. Civil registration should be accessible and provided at no cost.
Official data belong to the people and should be open to all. They should be open by default.
The data community should embrace the United Nations Fundamental Principles of Official Statistics as a starting point.
There is a need for governance and coordination of the data ecosystem.
African governments should acknowledge open data provided by credentialled data communities as acceptable sources of country statistical information.
Technology, new forms of data and other innovations should be actively embraced.
Data communities should promote a demand-driven data user culture spanning the entire ecosystem.
Privacy and intellectual property rights should be respected.
Data should be translated into information that is simple, understandable and relevant. Information must be timely, accurate, relevant and accessible.
Data must be driven by needs rather than for their own sake.
The data revolution in all its facets should be gender-sensitive.

Source: United Nations, Economic Commission for Africa, 2015.

Analysis of continental and national policy, legal and legislative frameworks, principles and protocols shows that the African data revolution can be understood in a two-dimensional sense:

- The considerable changes already under way within data ecosystems in Africa: the explosion in the volume, velocity and variety of data pro-

9 United Nations, Economic Commission for Africa, 2015.

duced within the continent, their dissemination, accessibility and use
- The data ecosystems to which the continent aspires: inclusive data ecosystems involving government, private sector, academia, civil society, local communities and development partners that tackle the informational aspects of development decision-making in coordinated ways; ones in which data are produced and used by a wide range of data communities and harnessed to impact development policymaking; and ones which embrace the core principles highlighted above. For example, the Government of Rwanda, which is one of the few African countries with an explicit Data Revolution Policy, follows the United Nations Expert Advisory Group and defines the data revolution as the "transformative actions needed to respond to the demands of a complex development agenda and private-sector needs, improvements in how data is produced and used; closing data gaps to prevent discrimination; building capacity and data literacy in "small data" and big data analytics; modernizing systems of data collection; liberating data to promote transparency and accountability; and developing new targets and indicators".

From the foregoing, it is apparent that the "African data revolution" implies something revolutionary both in what is already taking place within African data ecosystems and in what is aspired for or desired in ideal African data ecosystems.

1.2 What is revolutionary about the African data revolution?

From the general definitions of the data revolution to definitions that are specific to the African data revolution, the revolutionary aspect centres on at least five key factors:

(a) The exponential increase in the volume, types and speed of data available in African countries

While there are significant variations among them, many African countries are producing more, regular, and better-quality data, both from national statistical offices and other public agencies, but also from non-public-sector actors such as academic institutions, private-sector companies, development agencies, civil society organizations and citizen groups. For example, almost 9 out of 10 people in Africa now live in a country which has conducted a population census in the last 10 years. A third of all Africans live in a country where a census has been conducted since 2010, and almost all Africans (99 per cent) live in a country which has conducted a household survey in the last 10 years. Since 2005, 80 per cent of African countries have published a household survey which includes a health component.[10] At least 12 African countries have signed up to the Open Government Partnership, and at least 10 have established open data portals. In a number of countries, citizen groups and civil society actors are not only producing data that sometimes complement those produced by national statistical offices or fill in the gaps left by such data, but many of them are also promoting data-sharing and use among different stakeholder groups.

A number of data communities in Africa have either piloted or are experimenting with new sources of data collection or platforms for data accessibility and sharing. By tracking population movements through mobile phone data and by running analytics on social media, opt-in services such as the Kenya-based Ushahidi have made it possible for humanitarian organizations to better understand and respond more quickly and effectively to humanitarian crises and natural disasters. By looking at population flows and conducting real-time CCTV analytics to better understand traffic issues, some African cities, including Abidjan, Cape Town and Nairobi, seek to better plan for and accommodate their growing urban populations by building "smart cities" that have adequate health, water, electricity, education and public transport infrastructure. For example, Harvard University's Engineering Social Systems programme is overlapping mobile phone data with census data in Kenya to model the growth of slums, enabling city planners to better plan for and build public services such as water pumps and public toilets.[11] Through the Outthink Urban Planning Initiative, the city of Nairobi has partnered

10 Mo Ibrahim Foundation, 2016.
11 https://www.hsph.harvard.edu/ess/. Accessed 13 November 2016.

with IBM's Smarter Cities project to transform waste management.[12]

Work to harness the substantial increase in the volume, types and speed of data available is beginning to yield significant social, economic and scientific results. Over the last few years, a number of African countries have revised the methods and base year data used to calculate gross domestic product in order to reflect the structure and size of their economies more accurately. This has entailed accounting for economic activities previously omitted from national income accounting either because they were performed in the informal sector or because they occurred in sectors that were previously non-existent or underdeveloped such as information and communications technologies. The results of these exercises have been dramatic, and reflect the importance of high-quality data to national development. Nigeria's gross domestic product nearly doubled from $270 billion to $510 billion between 2013 and 2014, making it the largest economy in Africa. Kenya was catapulted from a low-income to lower-middle-income country, while Zambia's economy turned out to be 25 per cent larger than previously thought.[13] The rebased or updated figures not only provided better assessments of the size of the economy and the composition of and sectoral contributions to gross domestic product in a number of countries; the more accurate measures of these countries' economies provide an opportunity for policymakers and citizens to assess their fiscal circumstances better and prioritize their national development initiatives better. Besides the improved quality of official statistics, non-traditional sources of data collection, processing, dissemination and sharing such as opt-in and mobile surveys and open data platforms are generating constructive debates about data quality, timeliness and governance frameworks across the continent.

(b) Increased availability and use of new types of data as well as new uses of old or conventional types of data

The "revolution" in the African data revolution also centres on the increased availability of new types and sources of data as well as new applications of old or conventional types of data. Data are increasingly available on novel types of variables and on new activities, peoples or regions that were previously excluded from many conventional types and sources of data. Some of the new types or sources of data have already been highlighted (such as citizen-generated data and satellite or remote sensing data). Others include "big data" or large quantities of complex data (sets) which can be curated and analysed to derive insights and meaning. In popular balance, big data is often associated with either social media or business data, or both. However, this is misleading, since not only are there a range of different types of big data, but different types of big data sometimes require different analytical approaches. For example, administrative data or data collected by government agencies for the purposes of registration, transactions and record-keeping (such as data for the administration of tax systems, social programmes and regulations) can be a type of big data. Especially in developed countries, some of these administrative data now come in large data sets with tens of millions of distinct observations and huge numbers of covariates. The increased scale, dimensionality and lack of structure in some of the administrative data in developed countries have led some scholars to conclude that administrative data not only constitute a distinctive form of big data but are significantly underutilized because, among other things, they are not easily shared (among State and federal authorities, for example, or ministries and national statistical offices).[14] While the issues of scale, multidimensionality, lack of integration and coordination, and inadequate sharing and underutilization are all relevant to administrative data in many African countries, a general lack of administrative data is a bigger and more urgent problem on the continent.

The growing availability of large data sets that capture the behaviour of millions of individuals, or even the complexities of the galaxy, has yielded extraordinary scientific results, including in Africa. As Johan Fourie has argued, the applicability of the data revolution to the past is especially valuable in areas with limited written records such as sub-Sa-

12 IBM, 2016.
13 Mo Ibrahim Foundation, 2016.
14 Connelly and others, 2016.; Einay and Levin, 2013; Einay and Levin, 2014.

haran Africa. Fourie documents how a new generation of economists, geographers and historians are rewriting African history using not only thousands of individual archival records but also geographical, climatic and demographic projections into the distant past. A new African economic history has come of age, in large part thanks to the data revolution.

Scholars such as Fourie have used analysis of the height of African military recruits during colonial rule (as a proxy for living conditions) to document the evolution of living standards of African peoples during an era of unreliable data. These data have turned out to be useful not only as a tool to measure the level of living standards over time, but also to test the effects of different colonial policies on African populations. "The data revolution has not only offered new variables to measure past events, but it has been especially useful as a contemporary outcome variable. In the absence of data on regional African economies, for example, Michalopoulos and Papaioannou use light density at night obtained from satellite imaging to test the impact of precolonial ethnic institutions and the impact of the borders drawn during the Scramble for Africa on current economic performance".[15] In this case, the data revolution is not only helping to solve a fundamental problem in the study of African history and economics (i.e. a lack of data); it is also leading to innovative uses for old or conventional types of data. The decision by a growing number of African governments to make their traditional data (i.e. data produced by national statistical offices) open by default or accessible through open data portals is also leading to increasing accessibility of this type of data and in some cases, new uses of it by various stakeholder groups.

(c) Principles of data governance, especially the principle of inclusion or "Leaving no one behind"

What is "revolutionary" in the African data revolution also centres on the fundamental principles for governance of the production, dissemination, use and ownership of data. Many of these have been highlighted in box 1 above, and they include the principles of inclusion and openness. One of the most "revolutionary" principles of the African data revolution is the principle of inclusion or of leaving no one behind. At the heart of the principle is the idea that everyone matters or counts and that people must be counted or captured in formal data processes such as civil registration to allow the government to plan for their basic social, economic and other needs. In the context of the Sustainable Development Goals, this principle amounts to the idea that no goal, target or indicator should be considered met unless it has been met for every person around the globe.

Civil registration is particularly important as far this principle is concerned. Civil registration is the means by which countries keep continuous and complete records of births and deaths and the issue of birth and death certificates, which then dictate several formal civic, personal, professional, business and political activities and transactions. In many countries, it is through civil registration that individuals are admitted into schools and hospitals, gain nationality and formal employment, vote or present themselves for electoral office, buy and transfer properties and access financial and legal services. To be excluded from civil registration is in many cases synonymous with exclusion from public services. In spite of this, 46 African countries do not have complete civil registration systems to register births. By one estimate, 83 per cent of Africans live in a coun-

15 Fourie, 2015.

try without a complete and well-functioning birth registration system. Less than one in five births occur in a country with a complete birth registration system, while 87 per cent of deaths occur in countries without complete death registration systems.[16]

(d) Data-related rights and freedoms

What is "revolutionary" in the African data revolution also centres on data rights and freedoms. The Consensus is not explicit on this point, but underlying its many fundamental principles are data-related rights and freedoms. A number of African governments have policy, legal and legislative frameworks that either explicitly or implicitly provide for data-related rights and freedoms. These include:

- The right to be counted, as an integral part of the principle of leaving no one behind
- The right of access to information (for example, information held by the State – a right embodied in article 35 of the Constitution of Kenya and article 34 of the Constitution of Rwanda, as well as Rwanda's Access to Information Act)
- The right to participation (for example in data production, dissemination and use)
- The right to non-discrimination and equality
- The right to privacy and to ownership of personal data
- The right to freedom of expression (e.g. citizen-generated data)

The African data revolution, viewed in its aspirational dimension, is inclusive of all these rights and freedoms.

(e) A conceptual, ideological and developmental paradigmatic shift on the question of who and what officially counts, is counted, how, by whom and for whom

Underlying both the generic concept of the data revolution and the more specific notion of the African data revolution is a fundamental conceptual, ideological and developmental paradigmatic shift on the question of who and what officially counts, is counted, how, by whom, for whom and for what purposes. The principles of inclusion, data openness, timeliness, relevance and ownership all stand in contrast to the principles and paradigms that have governed the production, dissemination and usage of official statistics at some points in the African past, notably during the colonial period. What and who counts and is counted, by whom, how and for whom is more than just a matter of technical considerations or the applicability and utility of various statistical methods and techniques. At its heart, this question embodies political economy and ideological considerations of what are considered national priorities by governments in power or dominant interest groups in society. Accordingly, in the past official statistics in Africa have been used either as tools for social, economic and political exclusion and exploitation or in order to further inclusion, conflict resolution and integration, depending on prevailing dominant ideologies, political economies and development paradigms. The broader definitions of development and the principles of inclusion, participation and empowerment embedded in both the Sustainable Development Goals and the African Union's Agenda 2063 mark a fundamental shift in the conception of what and who counts and is counted, by whom and for whom. This report emphasizes that the question of who and what officially counts and or is counted, how, by whom and for whom is determined not just by technical considerations but by political, ideological, cultural and historical priorities or factors. That the African data revolution answers this question in fundamentally different ways from past behaviour and practices is part of what is revolutionary in the African data revolution.

1.3 Purpose of this report

This is the inaugural Africa Data Revolution Report, the first in a series of biennial reports that will focus on developments in national data ecosystems in Africa. It highlights the current state of the African data ecosystem, its desired end state, and the gaps in between. It also provides recommendations for creating data ecosystems capable of harnessing the full transformative potential of the data revolution in Africa. The specific objectives of this report are threefold:

- To review and map components of national data ecosystems in Africa, with a view to assess-

16 Mo Ibrahim Foundation, 2016.

ing their ability to help advance the attainment of national and regional development plans, as well as the Sustainable Development Goals
- To identify common systematic challenges that hamper the full realization of the African data revolution
- To put forward recommendations for accelerating progress towards the desired end state of the data ecosystem in Africa, in keeping with the aspirations expressed in the African Data Consensus.

1.4 Methodology

Data for this report were compiled using a combination of methods, including the following:

- In-depth country case studies in 10 African countries, aimed at understanding their national data ecosystems: key actors; legal, policy and legislative frameworks; infrastructure, technologies and tools; and the dynamic interactions among actors, subject to the limits imposed by legal, legislative and policy frameworks and infrastructural and technological limitations. In a majority of countries, country researchers worked closely with national statistical offices and with UNDP country offices.
- Review of scholarly, policy and grey literature on the data revolution globally and the African data revolution in particular. The literature reviews were conducted by the lead author and country researchers, with the latter taking the lead on each of their case study countries: Côte d'Ivoire, Ethiopia, Kenya, Madagascar, Nigeria, Rwanda, Senegal, South Africa, Swaziland and the United Republic of Tanzania.
- Interviews with key informants from representative data communities in each case study country (such as national statistical offices and other members of official statistics communities, representatives of the private sector, civil society, scientific and citizen-based data communities, as well as key players and stakeholders in open and big data communities).

1.5 Outline

The rest of this report is organized as follows. Section 2 maps the current data ecosystem in Africa, highlighting its strengths and limitations with specific reference to the requirements of the Sustainable Development Goals, Agenda 2063 and various national long-term development plans. Section 3 presents case studies of some innovations within various data communities in Africa, while section 4 concludes the report.

2. Data ecosystems in Africa

A data community is a group of people who share social, economic, political and/or professional interests in data across the entire data value chain – production, analysis, management, dissemination, use and storage. 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.[17] A data ecosystem encompasses multiple data communities (e.g. public, private and civil society actors); different types of data (old and new); institutions, laws and policy frameworks; technologies, platforms and tools; and the dynamic interactions among the actors within prevailing technological, infrastructural, legal, policy and other constraints. In this report, the term "data ecosystem" in Africa refers to the state of national data ecosystems in individual African States. The rest of this chapter examines the various data communities, institutions and frameworks, and technologies, platforms and tools, that constitute national data ecosystems in African countries.

Data communities can be categorized variously - for example, by sector, types of data, technological platforms, tools and methods. If classified by sector, for example, one might have public-sector-based data communities such as official statistical systems or national statistical systems, private-sector data communities, civil society data communities, scientific data communities and citizen-based data communities. If classified by types of data, in addition to the sectorally based classification, one might have open and big data communities. Each African country has a unique national data ecosystem depending on the distinctiveness of its data communities; types of data; institutional frameworks; data infrastructure, technologies, platforms and tools; and the dynamic interactions among the actors, which are subject to different technological, infrastructural, legal, sociopolitical and other constraints. Figure 2 below is a representation of the national data ecosystem in Swaziland.

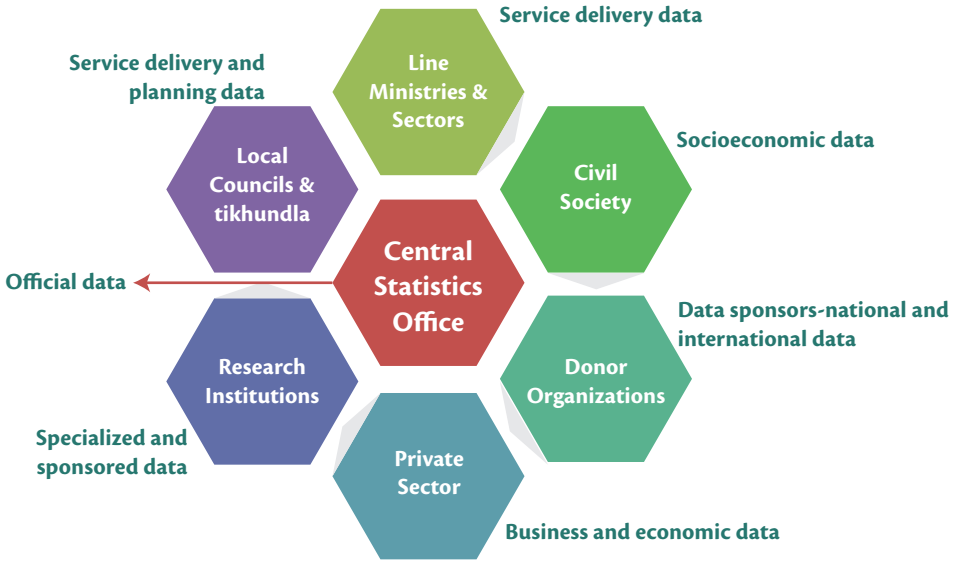
This report focuses on data communities as classified by sectors and types of data, as follows:

- Official statistics data communities
- Private-sector data communities
- Civil society sector data communities
- Scientific data communities
- Open data communities
- Big data communities
- Citizen-based data communities.

This classification is also adopted partly for ease of presentation. There can be significant overlaps between data communities. For example, open data communities can be inclusive of or interact with official statistics communities, citizen-based data

17 This definition draws from both the Open Data Charter (opendatacharter.net) and the Africa Data Consensus.

Figure 1: Data ecosystem map and data specialties and needs in Swaziland



Data Producers	Data Users	Infomediaries	Data Producers-Users	Data Sponsors
Central Statistical Office	Line Ministries	Central Statistical Office	Central Statistical Office	UN System
Line Ministries	Tikhundla	UN System	Line Ministries	African Development Bank
Local Councils	Local Councils	African Development Bank	Local Councils	European Union
Tikhundla	Civil Society	European Union	Tikhundla	Donor Organizations
Civil Society	UN System	Media/ICT/Technology	Civil Society	Private Sector
UN System	African Development Bank	FSE and CC	UN System	FSE and CC
African Development Bank	European Union		African Development Bank	
European Union	Donor Organizations		European Union	
Donor Organizations	Private Sector		Donor Organizations	
Private Sector	Media/ICT/Technology		Private Sector	

Source: Hlatshwayo, 2016.

communities and big data communities. As already mentioned, some official statistics (e.g. administrative data), as well as some scientific and citizen-generated data, could constitute types of big data. These overlaps also speak to some of the dynamic interactions between actors, varieties of data and technological platforms within national data ecosystems. Section 2.1 provides an overview of some of the significant transformations that have occurred within these national data ecosystems over the last decade. Subsequent sections examine specific data communities in detail.

2.1 Transformations in African national data ecosystems in recent decades

National data ecosystems in many African countries have undergone significant transformations over the last decade. These have included conceptual, legal, legislative, policy, technological, infrastructural and governance or management changes. These transformations have moved many national data ecosystems from a state of entrapment in a vicious cycle of limited data use, underfunding for statistics and underperformance of statistical services to the nascent stages of the African data revolution.

Until about a decade ago, national data ecosystems in most African countries were largely characterized by the predominant influence of the "official statistics community" or the national statistical system. Nevertheless, various types of data were being produced by non-State actors, who were mostly unacknowledged by the official system. For example, African universities and research institutes have long produced various types of scientific data, including data on climate, poverty, agriculture, health and the environment. Local and international non-governmental organizations and aid agencies such as Ox-

fam, Action Aid, Save the Children, World Vision and CARE have always produced project-based and advocacy-based data on food security, water and sanitation, governance, gender, human rights and natural and humanitarian disasters. Providers of financial services such as banks and insurance companies have been major players in the production, brokerage and use of various types of economic, trade and financial data, even if in many cases these data have always been proprietary. As section 2.2.7 below demonstrates, even citizen-generated data communities are not new to Africa.

What has changed in the profile of these non-State-based data communities compared to about 10 years ago is not limited to the exponential growth in the volume of data produced by them or the increasing speed or timeliness of the data. Rather, it includes the increasing realization by both official statistics and other data communities that, firstly, some of the data produced outside the official statistics data community can just be as high-quality, relevant and timely as official statistics, and secondly that some of these data fill or have the potential to fill key gaps in official statistics.

Until about 10 years ago, this willingness to consider the quality and relevance of data produced outside the official statistics community was rare in much of Africa. At any rate, legal, legislative and policy frameworks governing the production, dissemination and use of data did not entertain or anticipate these possibilities. The question of who and what officially counted and how this question was to be answered, was by law, the sole responsibility of the official statistics community.

In spite of its pole position within African data ecosystems, the national statistical system faced numerous challenges. In a foreword to Ben Kiregyera's book *The Emerging Data Revolution in Africa* Trevor Manuel, the former South African finance minister, summarizes the situation in most African countries until about a decade ago. "For a long time, the socio-economic-cultural environment in Africa was not conducive to production of good statistics. Effective demand for data was low, there was a gap between those who produce data – statisticians – and those who need and use data – policy and decision makers, politicians, planners, analysts and academia and programme managers; there was too much focus on collecting data required to meet immediate needs as opposed to building capacity for sustainable data production; infrastructure for statistical production, management and dissemination was not up to scratch".[18]

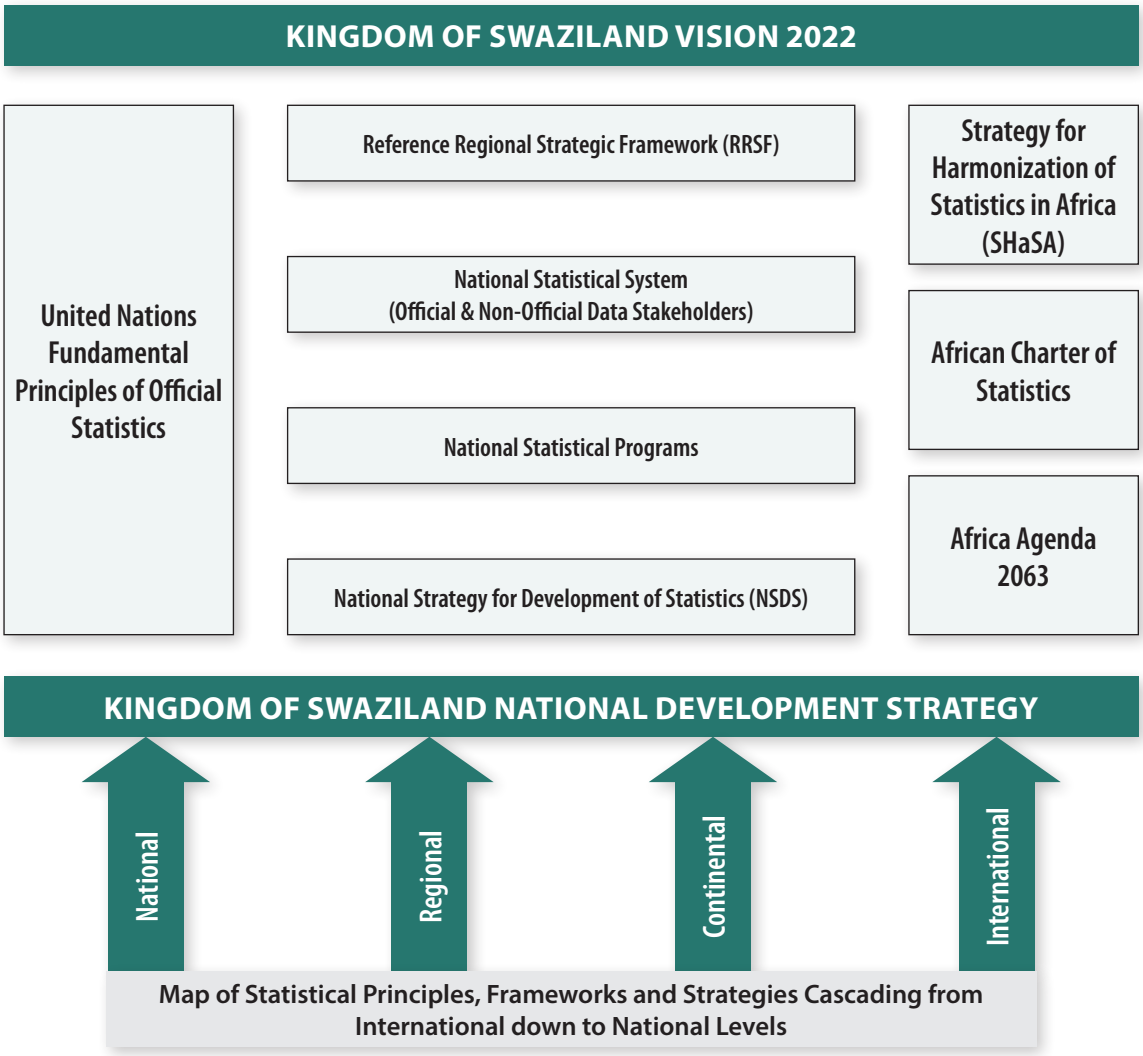
The national statistical system in most African countries was trapped in what has been called a vicious cycle of limited data use, underfunding and underperformance. Many national statistical offices were characterized by irregular data collection, a lack of financial and human resources, politicization or a lack of functional autonomy, and dominance of donor interests over national priorities.[19] The results of this were serious limitations on the scope, relevance, quantity and quality of data, and their limited use by policymakers and key stakeholder groups. This in turn contributed to limited investment in data infrastructure, technology, human and financial resources. As Krätke and Byiers have observed, "if incomplete or low-quality official statistics do not serve to meet the needs of policymakers, they in turn will have allocated less funding or resources towards the national statistical system, in turn lowering the demand and worsening the statistics produced".[20]

By one account, in many African countries during the 1980s and through the 1990s "statistics did not appear to be useful or contribute significantly in the arena of decision-making. They became supply-driven goods without a corresponding eager market demanding to make use of such products and services. For instance, in the case of Kenya, the census of the decade of the 1990s became a subject of political ridicule whilst the results of the census from Nigeria were not accepted until eight years later".[21] This vicious cycle started to change about a decade ago, culminating in what are now the nascent stages of the African data revolution. The changes that have ushered in the African data revolution include the following:

18 Kiregyera, 2015.
19 Ibid. See also Mo Ibrahim Foundation, 2016.
20 Krätke and Byiers, 2014.
21 Lehohla, 2008, quoted in Krätke and Byiers, 2014.

a) The explosion in the volume and types of data produced by multiple data communities, including data communities outside the official statistics community or the national statistical system. Increasingly many data are produced and used by actors outside the national statistical system, including scientific communities (e.g. remote sensing and satellite data); non-governmental and civil society organizations (e.g. aid, humanitarian, project-based and sector-based development data, especially in the areas of water, sanitation and health); private-sector data communities (e.g. financial, retail and minerals data); and citizen-generated and crowd-sourced data (e.g. citizen science and social media data).
b) The rise of open data policies and efforts to make official statistics open by default. As many as 12 African countries have now signed up to the Open Government Partnership, and at least 10 have created open data portals, where a number of official statistics can be accessed freely online.
c) The rise of big data and data mining not just by commercial companies but also by State agencies motivated by national security and other considerations. Countries with significant private sectors such as Egypt, Kenya, Nigeria, Senegal and South Africa are witnessing the collection, analysis and use of both social media and business data on a considerable scale, especially in the telecommunications, financial and retail service sectors.
d) Legal, legislative and policy reforms allowing for more integrated, timely, open, high-quality data and applied through a more inclusive, transparent and participatory process. These include constitutional reforms that guarantee the

Figure 2. Diagrammatic representation of international, regional and national influences on statistical principles and protocols in Swaziland



Source: Hlatshwayo, 2016.

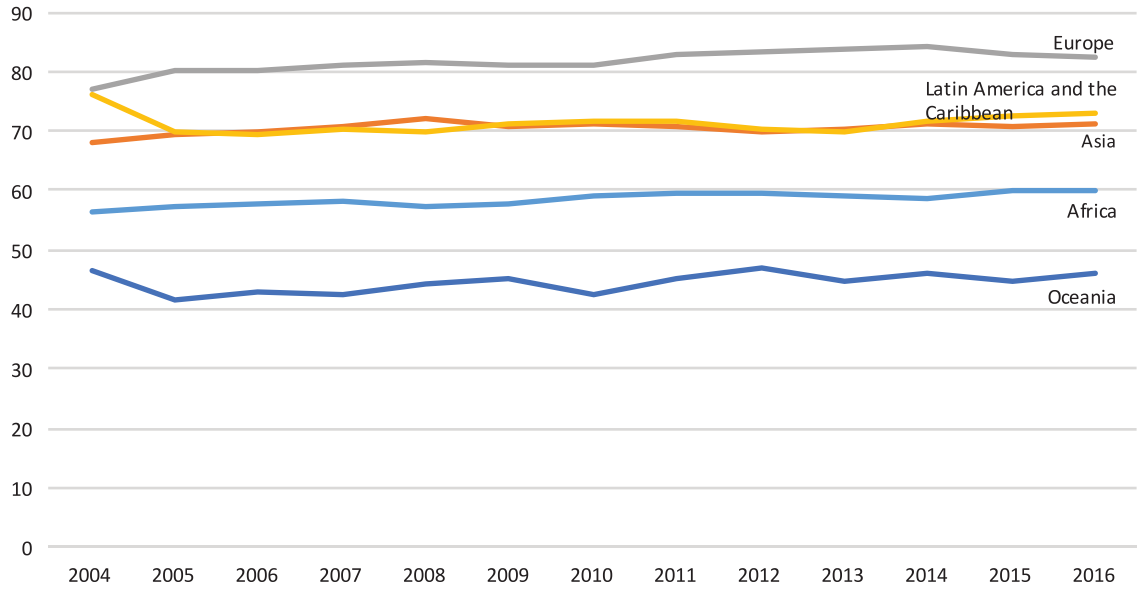
right of citizens to access information held by government agencies (as in Kenya and Rwanda); policies that seek to harness the data revolution for accelerated sustainable development; and reforms to statistics-related laws and legislation which are intended to enhance data quality, timeliness and usability (as in Ethiopia, Nigeria, Senegal and the United Republic of Tanzania).

e) The rise of new data technologies, platforms and infrastructure, including mobile and Internet-based technologies and the increasing use of computers and digitization.

f) Improvements in national statistical system capabilities. While Africa continues to have a lowaverage level of statistical capacity according to World Bank rankings (see figure 3 below), significant improvements have occurred in many African countries over the last decade. These changes have been driven by a number of factors. Firstly, the effort to measure progress towards the United Nations Millennium Development Goals led to significant investment in a number of national statistical offices in Africa. Secondly, the decision by African governments to sign up to a number of regional and global initiatives on improvement of statistical principles and practices, such as the Marrakesh Action Plan for Statistics (2004), the Reference Regional Strategic Framework for Statistical Capacity-Building in Africa (2006), the African Charter of Statistics (2009), the Strategy for the Harmonization of Statistics in Africa (2010) and the Busan Action Plan for Statistics (2011), helped to strengthen a number of national statistical systems. The decision by African governments, development partners and civil society to subscribe to the "managing for development results" framework, with its focus on the outcomes of development processes, coupled with increased democratization, transparency and accountability initiatives that focused on the use of data as a tool in the measurement of accountability and peer review, also helped. Figure 2 shows the many global, regional and national initiatives that have helped strengthen many national statistical systems in Africa, using the example of Swaziland.

In spite of these improvements, a number of challenges remain within the many national statistical systems in Africa. First, sub-Saharan Africa not only still has the lowest average level of statistical capacity in the world, but it also has the largest number of countries which have experienced a decline in capacity in the past 15 years - 11 out of 48 countries.[22] A recent study by the Mo Ibrahim Foundation reveals that up-to-date key indicators regarding employment and the labour market are

Figure 3: World Bank Statistical Capacity Indicator regional average (2004-2016)



Source: World Bank.

22 Krätke and Byiers, 2014.

unknown in more than half of African countries because they have not conducted a labour force survey in the past decade. Only half of the continent's population lives in a country that has carried out more than two comparable household surveys in the past 10 years. Changes in levels of poverty for half the continent's population are unknown, and only 29 per cent of countries have published a household survey with an education component since 2005. Only seven countries in Africa use the 2008 United Nations System of National Accounts, and less than a third of African countries have produced industrial data since 2006. Just over half of Africans live in a country which has conducted an agricultural census within the past decade. Information on the structure of the agricultural sector and landholders remains unknown for almost half of the continent's population.[23]

Secondly, a number of national statistical systems still suffer from a lack of functional autonomy and politicization of data. According to the Reference Regional Strategic Framework for Statistical Capacity-Building in Africa, in 2010 only 12 of the 54 members of the African Union were considered to have an autonomous national statistical office: Angola, Burkina Faso, Cabo Verde, Chad, Egypt, Ethiopia, Liberia, Mauritius, Mozambique, Rwanda, Uganda and the United Republic of Tanzania. Moreover, the Mo Ibrahim Foundation found that population, education and agricultural data in some African countries are often politicized for resource allocation purposes: "Population size may be used for budget allocations, or the allocation of parliamentary seats, or there may be political sensitivity around the number of ethnic or national minorities in the population. … There are often discrepancies between the enrolment rates shown in administrative and survey data, with the education enrolment rates stated in administrative data larger than those found in survey data. These divergences coincide with shifts toward top-down financing to education through per-pupil central government grants. … Agricultural censuses may be politically sensitive for reasons around national subsidies for agricultural input, or contain contradictory information around yield due to different statistical methods chosen to inflate these for political reasons."[24]

Thirdly, most national statistical systems are still underfunded and understaffed and lack critical data technologies, tools and infrastructure. According to the World Wide Web Foundation's Open Data Barometer, a majority of African countries can be classified as capacity-constrained: unable or unwilling to establish sustainable open data initiatives because of any number of factors, including Internet access and affordability and the capacities of public, private and civil society actors.[25]

In some countries, donors still provide up to 80 per cent of the budget of national statistical offices.[26] It is therefore not surprising that donor interests still prevail over national priorities in some countries. Finally, in most countries, there is still limited data-sharing, integration and coordination. There are still legislative and policy impediments to the production, dissemination and use of data other than those produced through the national statistical system, even where non-State actors have the capacity to produce and have produced high-quality data on activities, regions or persons previously excluded from conventional data.

All these examples provide evidence that the recent changes in African national statistical systems are far from sufficient or uniform across countries. Much needs to be done if the principle of leaving no one behind set out in the 2030 Agenda for Sustainable Development is to be met. That means improving not just the capacity of national statistical systems but also that of other data communities to collect, process, disseminate and use high-quality, timely, relevant and inclusive data. It also means facilitating greater interaction, coordination and integration among the multiple communities. A number of countries and national statistical systems recognize this and are taking steps to create national data ecosystem, in keeping with the aspiration embodied in the African data revolution. Section 2.2 takes a closer look at each of the key data communities in the current national data ecosystems in Africa.

23 Mo Ibrahim Foundation, 2016.
24 Ibid.
25 http://opendatabarometer.org/3rdedition/regional-report/africa/. Accessed on 1 November 2016.
26 Krätke and Byiers, 2014.

2.2 Key data communities within national data ecosystems in Africa

2.2.1 Official statistics communities or national statistical systems

There is no internationally agreed definition of official statistics, although various efforts have been made to develop common standards identifying the characteristics of official statistics, most notably through the Fundamental Principles of Official Statistics adopted by the United Nations General Assembly.[27] Most countries have legislation or policies that designate what constitutes official statistics. On the basis of a combination of international norms and the provisions of different national laws and policies, official statistics can be defined as a subset of statistics produced by official government or public agencies (such as national statistics organizations and relevant sectoral ministries) and/or specialized intergovernmental or international organizations or agencies, such as the World Bank, the International Monetary Fund or a variety of United Nations agencies. Official statistics typically include economic, social and demographic statistics (for example, national accounts, or statistics on industry, agriculture, trade, transport, tourism, population, health, education, labour and wages) produced by public agencies designated by law or policy to be the official producers or publishers of such statistics (in population and agricultural censuses, national household and demographic surveys, etc.). Such agencies might be either centralized statistical offices (e.g. national statistical offices or bureaus of statistics) or decentralized public agencies, in the case of agricultural or industrial statistics, among others.

A national statistical system is the ensemble of statistical organizations and units within a country that jointly collect, process and disseminate official statistics on behalf of the national government.[28] The national statistical system is also sometimes called the official statistics system, which can be conceived of as a government-wide system of policies, practices, processes, underlying data sources and people that are involved in producing and disseminating official statistics.

"Official data communities" means all the constituent members of the national statistical system or "all the actors and organizations in a country that collect, process and disseminate official statistics on behalf of the government".[29] As the sections below demonstrate, there are also a number of economic, social and demographic statistics produced by many non-State actors (e.g. researchers, the private sector, civil society and citizen groups), including those which challenge or fill in the gaps in official statistics. These statistics are not considered official statistics, and their producers are not considered part of the national statistical system.

Official statistics can be and have been used as a tool for social inclusion, integration and development, but equally for social exclusion, economic extraction, exploitation and political exclusion. While most national statistical systems in Africa are currently on a trajectory towards using official statistics as a tool for accelerating inclusion, integration and sustainable development, courtesy of global, regional and national initiatives such as the Sustainable Development Goals, Agenda 2063 and national long-term development plans and strategies, historically this has not always been the case.

Many African countries have constitutions, laws or specific legislation that provide for the production, dissemination and use of official statistics. These laws define official statistics. Who can produce them? How, and for whom? They also stipulate how official statistics may be disseminated or distributed. These are not just technical questions or determinations. They are inherently political and ideological questions, and different governing regimes have determined them variously, in the light of various prevailing political and ideological priorities and technical capabilities going back to the colonial period.

Modern national statistical systems in Africa emerged during the colonial period, and were

27 United Nations, Statistics Division, 2013. See also United Nations, Statistics Division, 2003.
28 Organisation for Economic Co-operation and Development (2002).
29 Krätke and Byiers, 2014.

primarily used by the colonial regimes "as tools for extracting revenues through taxation, whether to impose hegemony or integrate African countries into the monetary economy by turning them into fiscally viable States".[30] Activities that were not considered to be relevant or to contribute to these colonial objectives were not recorded or captured by the official colonial statistics. This included much of the activity in the informal sector, laying the groundwork for the exclusion of this sector from the bulk of official statistics, which continues to this day. This bias against the informal economy continued in the immediate post-independence period, courtesy of "modernization" and "industrialization" paradigms that prioritized the collection and publication of data on the industrial and large-scale agricultural sectors at the expense of informal manufacturing and smallholder agriculture. This bias was informed by the problematic assumption either that the informal sector would sooner or later be absorbed into the "formal sector", or that the informal sector did not meaningfully contribute to economic development. The relative technical difficulty of collecting data on the informal economy reinforced these biases, although the decision on what and who officially counted under colonial or post-colonial governments was not primarily driven by technical challenges. As Jerven and Lipton have shown, Asian and Latin American countries found solutions to this problem a long time ago.[31]

"Colonial administrations were originally established in order to draw on and extract resources of other countries for commercial purposes. The amount, location and potential value of these resources have affected the objectives and as such the production and usage of official statistics in African countries in several ways – both during colonialism and post-independence. These also shaped the production and usage of official statistics: the attitudes of colonizers towards public administration".[32]

Colonial legal and legislative regimes and political and economic ideologies simply did not take into account that much of the activity in the informal sector and many of the people engaged in it were not counted. However, it was not just informal activities that were excluded from official colonial statistics. Entire groups of peoples or regions could be included or excluded on the basis of colonial political and economic priorities and considerations. Krätke and Byiers have shown that colonial administrations not only favoured specific ethnic and social groups but that, through official statistics, they introduced demographic demarcations, or the frameworks for population statistics, that reflected these preferences. The legacy of this is still being felt in some African countries. As the recent study by the Mo Ibrahim Foundation found, population censuses in some African countries are often politicized.

In the immediate post-independence period (the 1960s and 1970s), many countries in Africa invested heavily in the production, dissemination and use of official statistics by building critical components of the national statistical system infrastructure. However, these efforts were severely disrupted by the structural adjustment programmes of the 1980s and 1990s. These programmes occasioned a significant decline in national statistical systems in many African countries, including the quality of the production of official statistics, their use and public investment in the system (in terms of human and financial resources, technology and infrastructure). By one account, during these decades "statistics did not appear to be useful or contribute significantly in the arena of decision-making. They became supply-driven goods without a corresponding eager market demanding to make use of such products and services".[33]

With the adoption of the Millennium Development Goals, as well as a number of international, regional and national statistical development initiatives (such as the Fundamental Principles of Official Statistics, the African Charter on Statistics and national strategies for the development of statistics), many national statistical systems in Africa have benefited from substantial reforms and investment since the early 2000s. These reforms have sought to strengthen not only the quality of statistics production but also their dissemination and use. They have included legal and legislative reforms geared towards using

30 Gervais and Marcoux, 1993. See also: Fisseha, 2012; Frankema and van Waijenburg. 2013; The Gambia, Central Statistics Department.

31 Jerven, 2013; Lipton, 2013.

32 Krätke and Byiers, 2014.

33 Lehohla, 2008.

official statistics as tools for promoting inclusion, integration and better decision-making. While many of these reforms still place national statistical offices at the heart of national statistical systems, and the latter at the heart of national data ecosystems, many of them also anticipate and provide for the participation of non-State actors within the national statistical systems and beyond. Reforms like these have taken place in countries including Côte d'Ivoire, Kenya, Mozambique, Nigeria, Rwanda,Senegal, South Africa and Uganda.[34]

Perhaps the most comprehensive legal and legislative reforms of this nature have occurred in Rwanda. Through a series of legal, legislative and policy reforms going back to 1997, Rwanda has strengthened the quality, use and dissemination of its official statistics; the human, financial and infrastructural capabilities of its national statistical system; and the coordination and governance frameworks for its national data ecosystem, including interactions between the official data communities and the scientific, private-sector, civil society and citizen-based data communities. The legal, legislative and policy frameworks that have made these changes possible include the following:

- Rwanda's Constitution (article 34) guarantees freedom of the press and information. Personal privacy is addressed in article 22, which provides that confidentiality of correspondence and communication shall not be subject to waiver except in circumstances and in accordance with procedures determined by law.
- The Access to Information Act, enacted in March 2013, provides that every person has the right to access information held by government bodies and certain private bodies "whose activities are in connection with the public interest, human rights and freedoms".
- Organic Law No. 45/2013 on the organization of statistical activities in Rwanda designates the National Institute of Statistics as the coordinator of the entire national statistical system. This act defines official statistics, as well as the actors involved in gathering them. It also legally compels persons and organizations to collaborate with surveys and share data gathered with the Institute. The Institute is designated as "the only institution entrusted to provide official statistics to the government, the business community and the public at large". However, the act provides that the Institute may allow other competent organs to provide official statistics once appropriate methods, quality and standards have been established and guaranteed.
- Rwanda's Penal Code (articles 286 and 287) creates criminal liability for the gathering and recording of personal information "likely to adversely affect the dignity or the privacy of people".
- Act No. 18/2010 relating to electronic messages, electronic signatures and electronic transactions recognizes the authority of any public institution to determine the mode of archiving, formatting and transmission and requirements for safe retention with respect to the acceptance, archiving and transfer of electronic documents. It also includes an obligation of confidentiality.
- The first phase of the National Strategy for the Development of Statistics (2009-2014) was designed to serve as a framework for the programmes, projects and activities of the national statistical system, with the aim of providing relevant, high-quality, timely and accessible statistics needed for evidence-based policymaking and decision-making in a more efficient and coherent manner. The second phase (2014-2019) aims at continuing the development of the national statistical system to respond to the statistical requirements anticipated for a typical middle-income country by 2020.[35]
- The national Data Revolution Policy (2016) provides for the harnessing of data, including conventional and non-conventional types of data produced both within and outside the national statistical system, for accelerated sustainable development in Rwanda.

The legal, legislative and policy reforms in Rwanda have been backed up by a high-level Presidential commitment, a national consensus on power-sharing, non-ethnic politics, a commitment to effective and equitable public service delivery and significant

34 See the country background studies prepared for this report, as follows: Nzayisenga, 2016; van Belle, 2016; Hlatshwayo, 2016; Mungai, 2016; Mdadila, 2016. Selassie, 2016; Rakotomanana, 2016.

35 Nzayisenga, 2016.

investment in the national statistical system. Investment in the production of official statistics increased from 90 million francs in 2002 to over 500 million francs in 2007.[36]

The results of these changes have been impressive. Since 2006, Rwanda has gradually improved its ranking in the World Bank's Statistical Capacity Indicator,[37] rising to ninth place in Africa in 2015, after Mauritius, Egypt, Morocco, South Africa, Tunisia, the United Republic of Tanzania, Senegal and Malawi. While considerable gaps and weaknesses remain, the country's national statistical system is able to collect and publish different types of data, including in areas relevant to the implementation and monitoring of the Sustainable Development Goals, as can be seen in table 1 below.

From the table it is apparent that the official statistics community in Rwanda consists of specialized national and sectoral public agencies under the overall leadership of the National Institute of Statistics. This is not unlike the set-up in many other African countries, although the legal and legislative frameworks in Rwanda provide for a more sharply focused and stronger vision for collaboration, engagement and coordination between the national statistical system and other data communities within the country, including academic, private-sector, civil society and citizen-based groups, a topic to which we will return below. For example, the United Republic of Tanzania's Statistics Act of 2015 defines official statistics as information produced, validated, compiled and disseminated by the Tanzania National Bureau of Statistics, government institutions and agencies. Section 20 (2) of the Act emphasizes that statistics produced by agencies shall qualify as official statistics if they meet the criteria and standards set by the Bureau and approved by its Director General. The Act insists on the need for provision of high-quality, reliable and timely official statistical information to the public. The Act also designates the Bureau as the coordinator of the national statistical system, entrusting it with establishing methods, standards and concepts and definitions for the production of official statistics, and also the power to regulate official information.

In spite of its impressive successes over the last decade, the national statistical system in Rwanda, like those elsewhere in Africa, continues to face a number of challenges. They include political, ideological and paradigmatic challenges related to the question of who and what officially counts, and to whom, and what is counted, by whom, how and for what purposes. There are still too many gaps in what and who is counted in Africa, which means that some people, issues and regions remain invisible to official statistics and the national statistical community. For example, adequate poverty data still do not exist for many African countries.

African national statistical systems also continue to face the challenges of inadequate financial, human, technical and infrastructural capacities. Many national statistical offices are still unwilling or unable to publish their data. Moreover, many continue to lack functional autonomy, and lack oversight over their own work plans and budgets. National statistical systems as a whole are still characterized by poor coordination, integration and management in many countries. This often leads to unnecessary duplications, overlaps or an inability to fill critical gaps by leveraging synergies and complementarities inherent within such systems.

All these challenges mean that while, as in Rwanda, many national statistical systems in Africa may have the ability to collect and publish data in key areas relevant to the Sustainable Development Goals, especially economic, social and demographic data, these data are in many cases not adequate to fully meet needs arising in the implementation and monitoring of the relevant Goals. In many such systems, aside from the social, economic and demographic data categories which benefited from significant investment during the era of the Millennium Development Goals, the situation is even worse. The data gap analysis of the United Republic of Tanzania (box 2) shows an example.

In sum, efforts are under way in most African countries to align national development plans with the Sustainable Development Goals where data collection for monitoring purposes is concerned. Explicit

36 Krätke and Byiers, 2014.

37 The Bank defines statistical capacity as a nation's ability to collect, analyse and disseminate high-quality data about its population and economy.

Table 1: Main sources of official data production in Rwanda

Data	Main source	Link
Land data	Rwanda Natural Resources Authority Rwanda Environment Management Authority	www.rnra.rw/ www.rema.gov.rw/
Soil data	National Institute of Statistics Ministry of Agriculture and Animal Resources Rwanda Agriculture Board National Agricultural Export Development Board Rwanda Natural Resources Authority Rwanda Environment Management Authority	http://www.statistics.gov.rw www.minagri.gov.rw/ www.rab.gov.rw/ www.naeb.gov.rw/ www.rnra.rw/ www.rema.gov.rw/
Climate data	Rwanda Meteorology Agency Ministry of Agriculture and Animal Resources	www.meteorwanda.gov.rw/ www.minagri.gov.rw/
Environmental data	Rwanda Natural Resources Authority Rwanda Environment Management Authority	www.rnra.rw/ www.rema.gov.rw/
Natural resources data	Rwanda Natural Resources Authority	www.rnra.rw/
Agricultural data	National Institute of Statistics Ministry of Agriculture and Animal Resources Rwanda Agriculture Board National Agricultural Export Development Board	http://www.statistics.gov.rw www.minagri.gov.rw/ www.rab.gov.rw www.naeb.gov.rw/
Population census data	National Institute of Statistics	http://www.statistics.gov.rw
Civil data (e.g. births and deaths, marriages and divorce)	National Institute of Statistics Ministry of Health Ministry of Local Government National Identification Agency	http://www.statistics.gov.rw www.moh.gov.rw/ www.minaloc.gov.rw/ www.nid.gov.rw/
Health data	National Institute of Statistics Ministry of Health	http://www.statistics.gov.rw www.moh.gov.rw/
Education enrolment and performance data (primary, secondary, tertiary)	National Institute of Statistics Ministry of Education Rwanda Education Board University of Rwanda Ministry of Local Government	http://www.statistics.gov.rw www.mineduc.gov.rw/ www.reb.rw/ www.ur.ac.rw/ www.minaloc.gov.rw/
Economic data	Ministry of Finance and Economic Planning National Institute of Statistics	www.minecofin.gov.rw/ http://www.statistics.gov.rw
Financial data	National Bank of Rwanda	www.bnr.rw/
Energy data	National Institute of Statistics Rwanda Energy Group Energy Utility Corporation Limited	http://www.statistics.gov.rw http://www.reg.rw/ http://www.eucl.reg.rw/
Transport data	National Institute of Statistics Rwanda Utility Regulation Authority	http://www.statistics.gov.rw www.rura.rw/
Water and sanitation data	National Institute of Statistics Water and Sanitation Corporation Limited	http://www.statistics.gov.rw www.wasac.rw/
Crime data	National Institute of Statistics Ministry of Justice Rwanda National Police	http://www.statistics.gov.rw www.minijust.gov.rw/ www.police.gov.rw/
Governance data (indicators)	Rwanda Governance Board	www.rgb.rw/
Election data	National Electoral Commission	www.comelena.gov.rw/

efforts to this effect have commenced in Côte d'Ivoire, Kenya, Nigeria, Senegal, South Africa and the United Republic of Tanzania. As a result of earlier investment in the production of data to monitor progress towards the Millennium Development Goals, a number of national statistical systems in Africa have gained some capacity to produce data that can be used to capture progress towards what might be called the social and economic development dimension of the Sustainable Development Goals (i.e. Goals 1, 2, 3, 4, 6, 7, 8, 9 and 10). With the exception of South Africa, whose national statistical capacity is among the strongest on the continent, in general the production, dissemination and

use of data relevant to Sustainable Development Goals 11, 12, 13, 14 and 15 remain relatively weak in a number of African countries.

What this means is that significant investment is needed to strengthen the human, institutional and technological capacities of most national statistical systems in Africa if the continent is to harness the data revolution for accelerated sustainable development. As these systems are already the strongest and best funded and equipped elements of the national data ecosystem in many African countries, a greater level of investment effort is required to strengthen the capacity, quality, accessibility and use of the other parts of national data ecosystems, including most notably data generated by citizens, civil society and the private sector and the corresponding data communities. Additionally, there is an enormous gap in data use, integration or interoperability – the ability to make the most of all the data that are already available within national data ecosystems, regardless of their sources of production, and the ability to integrate and/or triangulate the various types and sources of data generated within countries.

2.2.2 Private-sector data communities

The rise of big data and the open data movement, together with advances in technology, have recently shed light on the considerable amounts of data that the private sector produces in Africa and the poten-

Box 2: Data requirements in pursuit of the Sustainable Development Goals and gap analysis in the United Republic of Tanzania

In view of the possible clustering of Sustainable Development Goals indicators, a categorization into economic development goals, social development goals, environmental sustainability and "enablers of sustainable development outcomes" is applied for this report. The gap analysis suggests the following:

Economic development goals (Sustainable Development Goals 1, 2, 7, 8, 9 and 10)
Data are available for most indicators under this category. Goals 1 and 2 are mainly a continuation of the Millennium Development Goals, and hence a monitoring framework already exists. Proper scheduling of household surveys, such as household budget surveys and national panel surveys, will be ideal to ensure a regular data flow, especially in tracking poverty in the country. However, comparison of statistics across these data sources will depend on consistency in survey methods used, entailing a need for harmonization of survey methodologies. Establishment of new data collection systems may be required, principally for eight indicators: two for Sustainable Development Goal 8, four for Sustainable Development Goal 9 and two for Sustainable Development Goal 10 . Analysis for this report was unable to suggest any possible source of data for these indicators. Further, this cluster will comprise four indicators to be monitored globally: one for Sustainable Development Goal 1 and three for Sustainable Development Goal 10.

Social development (Sustainable Development Goals 3, 4 and 6)
Data are available for most of the indicators in the country. The challenge is the frequency of updating and the level of disaggregation. Monitoring of Sustainable Development Goal 6 will be critical. The analysis for this report suggests that seven indicators for Sustainable Development Goal 6 will require new data collection systems. Monitoring of this cluster will be at the local level, as there are no global indicators.

Environmental sustainability (Sustainable Development Goals 11, 12, 13, 14 and 15)
This category has introduced indicators which require data beyond what is traditionally collected in the country's national statistical system. More capacity development around these indicators in terms of equipment, skills and manpower is thus key for effective monitoring of this dimension. A total of 27 indicators will require new data collection systems. Sustainable Development Goal 15 will be critical in terms of availability of data. The analysis shows that 11 indicators for this goal will require the establishment of new data collection systems.

Reporting on the environment has remained relatively weak compared to other spheres covered by the Sustainable Development Goals. However, the United Republic of Tanzania presents an opportunity to improve on environmental reporting, particularly through the use of the State of Environment Report. This report, required under the Environmental Management Act, describes the state of the environment using the DPSIR (Drivers, Pressures, State, Impacts and Response) framework of analysis.

Enablers of sustainable development outcomes (Sustainable Development Goals 16 and 17)
The goals in this category aim to create an environment for achieving other Sustainable Development Goals. Data are available for most of the indicators. Further, most of indicators for Sustainable Development Goal 17 will be monitored at the global level. It will be the responsibility of the national statistical system, coordinated by the National Bureau of Statistics, to provide data on 41 indicators to the global monitoring system. Environmental indicators will need much of the attention, mainly in terms of capacity to collect data, set standards and analyse data.

Source: Mdadila, 2016.

tial use of such data for sustainable development if harnessed effectively. While these changes mark a significant transformation in the volume, scope, variety, speed and availability of data generated by the private sector , this sector has always been a major producer of data in Africa, even if this fact has not often been shared or widely acknowledged. Financial services, oil, gas and mining companies and property, construction, industrial and agricultural companies have often produced significant amounts of economic, trade, environmental and consumer data. Typically, much of the data produced by the private sector has been deemed commercially sensitive and used mostly internally. This has long suppressed the profile of private-sector data communities within many national data ecosystems in Africa.

However, even before the data revolution, some private-sector companies, especially in the financial services subsector, often produced financial data on a regular basis. They include national stock or security exchanges (in Côte d'Ivoire, Egypt, Kenya, Morocco, Namibia, Nigeria, South Africa and Zimbabwe); national offices of transnational banks (e.g. Barclays and Standard and Chartered); international consulting firms (e.g. PWC, Deloitte, Ernst and Young); oil, gas and mining companies; and private-sector-based publications such as those of the Economist Intelligence Unit and *Africa Confidential*. With the rise of big and open data and the growth in the ICT sector, the private sector is increasingly becoming a critical and dynamic player within African data ecosystems. Typically, these companies have sophisticated information systems and data mining and analytical capabilities and collect enormous amounts of data.

The laws governing data production, dissemination and use in many African countries now anticipate and allow for the participation of non-State actors, including the private sector, within national statistical systems, subject to certain quality requirements and other conditions. While formal collaboration between private-sector and official data communities remains preliminary or marginal in most African countries, and while most data produced by the private sector are still restricted to internal use because of sensitivity concerns, the private sector is nevertheless increasing its engagement within African data ecosystems in several ways. Besides the growth of companies that share non-proprietary data with the public, there is considerable experimentation and innovation in data production, brokerage and use within the private-sector data community in Africa. Much of this experimentation has occurred in the areas of big data and technological platforms for facilitating crowdsourcing or citizen-generated data.

The Orange Data for Development Challenge in Senegal, run jointly since 2012 by the University of Louvain in Belgium and the Massachusetts Institute of Technology, is indicative of such experimentation. Participants were invited to find innovative ways to address societal development and welfare by harnessing mobile-phone-based data. Over 260 international research laboratories, including 11 in Senegal, were given access to anonymous data samples extracted from the mobile network in 2013. The challenge was focused on five national development priority areas, using questions designed by government ministries in the fields of health, agriculture, transport/urban planning, energy and national statistics. More than 150 teams worked on the data, and up to 60 high-quality projects were submitted to the competition. Out of these, 40 per cent were aimed at improving transport and urban development in Senegal, and 20 per cent and 15 per cent were devoted to health and national statistics respectively. The remaining 25 per cent were spread between agriculture, energy and data visualization and anonymization.[38] A similar Challenge in Côte d'Ivoire produced over 80 research papers covering diverse topics ranging from optimizing bus routes and analysing social divisions to studying disease containment policies.[39]

Elsewhere on the continent, private-sector firms, especially technology companies, are engaging with national data ecosystems by facilitating platforms for crowdsourced or citizen-generated data. These include the Google Live Traffic Alerts project in Kenya. In April 2016, Google began providing live traffic alerts in Kenya through its Google Maps mobile app when it is set to navigation mode. The app

38 Orange.com, Orange Data for Development Challenge in Senegal. Available from http://d4d.orange.com/content/download/43330/405662/version/3/file/D4Dchallenge_leaflet_A4_V2Eweblite.pdf.
39 de Montjoye, 2014.

provides updated information on whether congestion is expected, and how long drivers may be stuck in traffic using a particular route. It then suggests alternative routes, including explanations of their advantages. The service is based on crowdsourced map data and traffic information from its active users.[40] In South Africa, a few commercial infomediaries are consolidating and republishing some data held by the private sector, especially economic and financial data. They include EasyData, EconoStatistics, Quantec and TradingEconomics, which publish South African economic data, including financial, trade, microeconomic and macroeconomic data.[41] These infomediaries combine data from official statistics with data from various other sources. Their presence speaks to the importance of data-sharing and integration, but also points towards potential models for making some private-sector data available to broader groups.

While many of these examples are still small-scale pilot projects, they nevertheless highlight the potentially transformative power of the full utilization of the data held by the private sector in Africa. The challenge remains how to make these often proprietary data available for social or public purposes in ways that do not diminish their commercial viability for their producers. A variety of models for achieving this are currently being piloted in Africa and elsewhere. These include the Open Algorithm Project, a pilot platform for unleashing the power of big data held by private companies for the public or social good in a privacy-preserving, commercially sensible, stable, scalable and sustainable manner. It consists of an open technology platform and open algorithms running directly on the servers of partner companies, behind their firewalls, to extract key development indicators of relevance for a wide range of potential users, including national statistical offices, ministries, civil society organizations and media organizations. It is currently being piloted in selected African, Asian and Latin American countries with the support of a large number of organizations including Orange, the Massachusetts Institute of Technology Media Lab, Data-Pop Alliance, Imperial College London, the World Economic Forum and the World Bank.

The potential applications of private-sector-held data in Africa are enormous, and much experimentation is under way on how best to make high-quality relevant data easily accessible in a timely manner. Private-sector data communities have the potential to contribute significantly to the production, dissemination and use of data relevant to the Sustainable Development Goals. However, while most national statistical systems have been reoriented to address the data needs of the Sustainable Development Goals, no systematic effort has as yet been made to harness data held by private-sector communities for sustainable development in Africa either on a large scale or sustainably. Collaboration between the private sector and national statistical systems will be very important and necessary for the implementation of the Sustainable Development Goals. However, legal, technological, financial and human resource challenges to such collaboration remain.

2.2.3 Civil society data communities

Civil society data communities have long existed in Africa, even if the social, economic, political, environmental and other potential of data from this source has only recently been brought into sharp focus by the data revolution. Much of the data on poor and marginalized communities and regions, including data on water and sanitation, health and education and natural and humanitarian disasters has also often come from local, national and international civil society organizations. Because of the significant role that civil society has played in democratic and governance reforms in Africa, the data community in this sector also accounts for a considerable share of the data on human rights, gender, electoral issues and corruption produced and used on the continent. These include not just flagship reports such as Transparency International's National Corruption Index but also electoral and human rights data held by various local, national and international civil society organizations monitoring elections and human rights.

Civil society in Africa is vast and dynamic. It includes everything from international, national and local non-governmental and community-based organizations to the press and professional associations and groups, including most notably law

40 Mungai, 2016. See also Klosowski, 2016; Osamuyi, 2016.
41 van Belle, 2016.

Box 3. Civil society data communities in South Africa*

South Africa has a hugely active non-profit/NGO sector, with more than 85,000 non-profit organizations registered by the Department of Social Development, i.e. a ratio of one for every 60 citizens. As with for-profit organizations, most of them are very small operations. However, a number of them play a strong national research and advocacy role and generate valuable data directly or indirectly related to the Sustainable Development Goals. In a number of instances, they provide important alternatives to existing data, or estimates of data that are otherwise not available.

Although most non-profit organizations are unlikely to generate census-type data of the quality and scope which Statistics South Africa provides, they provide a valuable contribution in respect of marginalized groups such as children, the disabled and the homeless and hard-to-measure, unobtainable or unreliable statistical data on crime, violence, injustice, corruption and health. A number attempt to conduct regular and representative surveys, and pay specific attention to and report on methodological concerns.

The Early Childhood Review case study
Finding detailed and comparable statistics about vulnerable population groups is very challenging. Given South Africa's young population, investment in early childhood development is seen as an imperative by the United Nations and the national Government. However, disaggregated statistics on children under six are hard to access. The non-profit organization Ilifa Labantwana, the Children's Institute at the University of Cape Town and the Department for Planning, Monitoring and Evaluation collaborated on the South African Early Childhood Review 2016,** which integrates data from a huge variety of official and non-official sources to provide 40 indicators relating to maternal and child health, nutrition, support for care-givers and social services/grants. Many of the indicators are disaggregated by province.

* This section draws heavily from van Belle, 2016.
** Available from http://ilifalabantwana.co.za/wp-content/uploads/2016/05/SA-ECD-Review-2016-low-res-for-web.pdf.

Table 2: A sample of civil society organizations in South Africa producing data relevant to the Sustainable Development Goals

Christian AIDS Bureau for Southern Africa	www.cabsa.org.za	HIV/AIDS data and research
Centre for Conflict Resolution	www.ccr.org.za	Mostly research reports
Centre for the Study of Violence and Reconciliation	csvr.org.za	Research and surveys
Children Count	www.childrencount.org.za	40 Indicators on children in South Africa including demographics, income/grants, housing, education, health and nutrition
Corruption Watch	www.corruptionwatch.org.za	A very active non-profit organization tracking, documenting and fighting corruption in South Africa
Institute for Justice and Reconciliation	www.ijr.org.za	Research and policy in relation to justice and reconciliation in South Africa
Institute for Security Studies Crime and Justice Hub	issafrica.org/crimehub	Has maps on public violence and reported crime, and many other publications
Health Systems Trust	www.hst.org.za	Provides resources and information on health systems and primary health care in southern Africa, including detailed reports based on data from its own and various outside sources
Labour Market Intelligence Partnership	www.lmip.org.za	Research and data on skills and human resources development in South Africa
South African Institute of International Affairs	www.saiia.org.za	Think tank; carries out research on international issues
South African Institute of Race Relations	irr.org.za	A plethora of statistics on assets and income, business infrastructure, communications, crime and security, demographics, the economy, education, health, industrial relations and living conditions

societies, trade unions, environmental and other single-issue or multiple-issue advocacy and activist groups. While the depth, quality and availability of data generated by civil society may be questionable in some cases, the scope of the data they produce, as well as their relevance and timeliness to the Sustainable Development Goals and other national and or international development priorities, are considerable. In many countries, civil society data communities have always pushed to expand what and

who officially counts and is counted. They generally tend to subscribe to the principles of inclusion and participation. Their ability and willingness to work in relatively remote areas and with marginalized or difficult-to-reach populations has always meant that they have typically, albeit informally, filled key gaps in many official statistics in Africa.

With the new spotlight that the data revolution has shed on civil society data, and with the introduction of legal and legislative reforms that increasingly allow space for the contribution of multiple data communities in the production of official statistics, civil society data communities potentially have growing opportunities to contribute to the harnessing of data for accelerated sustainable development in Africa. Indeed, these communities have been very active players in the emerging data ecosystem in Africa. Not only have they helped push the legal and legislative reforms that are opening up national data ecosystems, but many of them have also been active players in open data platforms and big data experimentation.

But challenges remain. One relates to scale. Can civil society produce data that are relevant and timely at scale? Secondly, given the potential scope for non-State actors producing high-quality statistics to play a role within some national statistical systems as potential sources of official statistics, can civil society data communities meet the quality standards desired by the official statistics data community? Thirdly, can civil society data communities align their data priorities with the needs of the Sustainable Development Goals?

The civil society data community in South Africa (box 3) exemplifies some of the opportunities facing the civil society data community in Africa.

2.2.4 Academic or scientific data communities

Academic and research communities in Africa, which include universities, specialized research institutes, think tanks and individual researchers, are key producers of scientific data. Unlike the other data communities, these communities have not received much attention for their potential contribution to harnessing the data revolution for accelerated sustainable development. This perhaps reflects a number of past characteristics of this sector. Not only has it had limited engagement with many national statistical systems in the past, but Africa's scientific output compared to the rest of the world has been very low, and much of the scientific data and output from the academic community is often not readily available in user-friendly forms to other stakeholders. Sub-Saharan Africa accounts for less than 1 per cent of the world's research output. Much of Africa's research is either published in unindexed journals or remains unpublished as dissertations on university library shelves.[42] Moreover, Africa has fewer researchers than any other region of the world. Africa has about 35 scientists and engineers per million inhabitants, compared with 130 for India, 168 for Brazil, 450 for China, 2,457 for Europe and 4,103 for the United States. At 7 per cent, Africa has the world's lowest tertiary enrolment rate, against a global average of 30 per cent. A continent with about 12 per cent of the global population produces about the same amount of research output as the Netherlands.[43]

While these figures indicate that the African scientific data community is relatively small, the volume, quality and potential applications of the data produced by this community should not be underrated. More importantly, like other data communities on the continent, the scientific data community in Africa has undergone significant transformation over the last decade. Among the most notable phenomena has been the improvement in both the quality and quantity of scientific outputs from the continent. Sub-Saharan Africa (excluding South Africa and North Africa) more than doubled its yearly research output from 2003 to 2012. The region's share of global research increased from 0.44 per cent to 0.72 per cent over the same period. Between 2008 and 2014, the number of research articles published in Africa rose by 60 per cent, with Africa's share of global publications increasing from 2 per cent to 2.6 per cent. The number of researchers in Africa grew from 150,000 in 2009 to 190,000 in 2013, while the continent's gross expenditure on research and development grew from $12.9 billion in 2007

42 Francescon, 2017.

43 These figures are for sub-Saharan Africa only, excluding South Africa and North Africa. When South Africa and much of North Africa are added to the mix, Africa performs much better than this – see Ochieng, 2014.

Box 4: The scientific and academic data community in South Africa

South Africa is the greatest producer of academic research in Africa. This research is generated by a cohort of roughly 12, 500 academics and about as many researchers based at research councils as well as research institutes, some of them in the private sector. South Africa currently has 26 universities, but just over half of them are focused entirely (8) or partly (6) on vocational education. Despite ongoing efforts to build research capacity at the historically disadvantaged universities, most of the research still happens at the leading research universities. The top research universities have several dedicated research units or institutes, usually specializing in key disciplinary areas or sectors such as health, energy, astronomy, economics and demography. Few if any of these produce original, large data sets, but many produce excellent "snapshot" reports based on survey data which are often available online or on request. The top research universities have created institutional repositories which, while they make research reports and theses available, do not generally host data sets. This situation is expected to improve as the open access paradigm becomes more established, and as researchers discover the (citation and impact) value of opening up their primary data sources along with the research reports. However, even if they are made available, data consistency, quality, lack of metadata, lack of longitudinal series, representativeness and similar issues are likely to remain obstacles for many years to come.

The better-known university-based institutes that produce regular data sets include DataFirst at the University of Cape Town, the Southern Africa Labour and Development Research Unit, the University of Stellenbosch and the University of KwaZulu Natal. DataFirst is probably the leading research unit which preserves South African data sets and makes them available online to researchers. Currently, it hosts 329 data sets, with a heavy focus on socioeconomic survey data. DataFirst provides independent data services to researchers across Africa. Its primary role is that of an infomediary which collects and curates South African and African survey and administrative micro-record data sets, and provides infrastructure to researchers for easy access to these data sets. It offers a mixture of open data sets which are publicly available and downloadable from its data portal, licensed data sets where access is granted on individual request, and restricted data sets which typically require on-site access through a laboratory with secured access, because the data are too confidential to be released (the date are then analysed and processed on DataFirst's in-house secure data server). The public data portal holds 340 data sets, with a bias towards demographic and economic data.

In addition to providing access to data, DataFirst plays an active role in interrogating and improving data quality by conducting research into the methodologies used, analysing the metadata, and often harmonizing data, for example to make economic time series data such as the post-apartheid labour market series consistent. Ensuring that data are of high quality and usable and come with detailed metadata is an overarching concern. In some cases, DataFirst curates the primary data collected by associated research institutes or research programmes. It has also partnered with other African data producers in Kenya, Nigeria, Rwanda and the United Republic of Tanzania to curate their micro-data, under a programme known as the Accelerated Data Programme. Additionally, DataFirst offers workshops aimed at training researchers in the use and analysis of data sets and how to assess and ensure data quality in statistics. It publishes a series of technical papers and reports on data quality issues and selected data sets.

The Southern Africa Labour and Development Research Unit is known for its National Income Dynamics Study, a large-scale panel survey producing detailed, in-depth reports on changes in national income distribution and the underlying dynamics. So far four "waves" of the survey have been executed, spanning 2008 to 2015. At the University of Stellenbosch, the Bureau for Economic Research, the unit for research on socioeconomic policy and the Institute for Futures Research make a large number of research reports available (not all are free), although they do not publish the underlying data they collect. At the University of KwaZulu-Natal, the School of Development Studies makes some of its data sets available (also through DataFirst). There are also a number of other quasi-governmental agencies or organizations which, in the pursuit of their specific mandates, collect and produce primary data.

South African governments have, over the years, created eight largely independent research councils dedicated to specific areas of scientific enquiry. Apart from government-commissioned research – the primary reason for their establishment – they also undertake contract research from or for their particular industry. Examples are the Council for Scientific and Industrial Research, the Human Sciences Research Council, the Medical Research Council, the Council for Geosciences, the Agricultural Research Council and the South African Bureau of Standards. Each has its own in-house researchers, but will often outsource to or collaborate with university academics for specific research projects. All of them make much of their research output available, while a number of the largest ones have also published some of their data sets. The National Research Foundation is the vehicle through which the Government promotes the development of research capacity, since it is the major provider of research funding.

to $19.9 billion in 2013, or from 0.36 per cent to 0.45 per cent of gross domestic product.[44] In other words, the scientific data community is an integral and growing part of the data ecosystem in Africa.

The scientific data community produces various types of data in Africa. They include geospatial data, economic and demographic data, environmental and climatic data, and sectorally based data in sec-

44 Blom, Lan and Adil, 2016. See also United Nations Educational, Scientific and Cultural Organization, 2016, and Ochieng, 2016.

tors such as energy, water and health. These scientific data sets can be combined with existing data sets already being produced in national statistical offices to provide insights that neither source can provide alone. For example, given the lack of weather stations on the ground and the paucity of historical records in Africa, climate scientists combine meteorological observations from different sources to produce "gridded climate data" which can provide valuable information on the spatial and temporal pattern of climate variation. This data set can be linked to socioeconomic data sets from a census or demographic health survey to allow integrated analyses of the influence of climate on health, population dynamics and other socioeconomic characteristics of a society.[45] Moreover, the scientific data community is not just a producer and consumer of different types of data. It also plays a critical role as a broker of data within the ecosystem. According to a recent study, in 2012, 79 per cent, 70 per cent and 45 per cent of all research in southern Africa, East Africa and West and Central Africa, respectively, was produced through international collaboration.[46] Through their open data research networks, they also help advance the open data movement. As is demonstrated by the 11 research laboratories in Senegal that participated in the big data challenge, the scientific data community is also a key player in big data experimentation on the continent.

South Africa, the continent's scientific leader, best illustrates the strengths and limitations of the African scientific data community, with specific reference to the Sustainable Development Goals (box 4).

The scientific data community in Africa faces its own challenges. One of these relates to the question of relevance and timeliness of data. Scientific research cycles and policy data needs are not often aligned. Scientific research tends to take time, while most policy data needs occur in real time. Scientific research and data can also be highly technical and therefore relatively inaccessible to non-specialist data audiences. In some cases, brokers or interpreters may be needed to process and present such data in formats that are easier to use for different data communities. These challenges are common to most scientific data communities globally. The challenges that are unique to African scientific data communities relate to a lack of adequate financial and human resources; inadequate technological infrastructure, tools and platforms (such as a lack of indexed journals and online platforms, limited digitization of scientific publications, etc.); infrequent use of scientific data by the policy community; and a lack of coordination and cooperation with the national statistical system. In order to harness the full potential of data produced and held by scientific communities, Africa countries will need to address these challenges.

2.2.5 Open data communities

a) Open data

Open data are data that are freely available online and can be used, reused and redistributed or republished by anyone. While the most widely used sources of open data have so far been governments and government agencies (i.e. open government data), the private sector, civil society, academia, citizen groups and social media can be, and are, all sources of open data (e.g. open extractives, open aid, open contracting, open data science, etc.). Open data are characterized by three main attributes: availability and access (the data must be available as a whole and/or in a convenient and modifiable form); reuse and redistribution (they must be provided under legal conditions such as open licences that allow reuse, republication and remixing with other data sets); and universal participation (allowing use, reuse and redistribution by anyone for any purposes, including commercial and non-commercial purposes).[47]

It has been argued that open data can be used to create social, economic and business value by facilitating better governance, public services and decision-making, supporting new businesses and improving the climate for foreign investment. The McKinsey Global Institute estimates that open data used in seven sectors (education, transportation, consumer products, electricity, oil and gas, health

45 Grace and others, 2012; Pricope and others, 2013.
46 Blom and others, 2016.
47 World Bank, 2016; Davies, Perini and Alonso, 2013; Jesuit Hakimaki Centre, 2014; Open Knowledge International, 2016.

care and consumer finance) could create $3-5 trillion annually in economic value worldwide.[48] The World Bank estimates the direct annual economic value of open government data at about 40 billion euros throughout the European Union and nearly £2 billion in the United Kingdom.[49] The World Wide Web Foundation's programme on Exploring the Emerging Impacts of Open Data in Developing Countries sums up the potential of open data in development thus: "the open data movement holds out the promise of improving transparency, accountability, citizen participation and economic opportunity across developing countries. Citizens in Brazil, Nepal, and Nigeria can use publicly available data on government budgets to track and fight corruption, or to critique public spending policies. Developers and entrepreneurs across Latin America, Africa, and Asia can create web and mobile applications using government data on education, health, and crime, with the potential to promote smarter and more efficient local public services. And donors and advocacy organizations are investing in open data, opening their own data sets, or pushing for open data as part of open government reforms".[50]

In spite of these claims, there is as yet limited evidence that open data initiatives are delivering on these great promises in developing countries.[51] Firstly, the open data movement is relatively young. In its current form, it can be dated to about 2010, meaning that most open data initiatives in the developing world are still in their infancy. There are still few if any large-scale rigorous studies of the impacts of open data on social, economic and other types of development. Much of the impact analysis is still ad hoc, focuses on isolated case studies or is anecdotal. Much remains unknown about the relationship between socioeconomic and political contexts, open licences, technical platforms and the dynamics of data use and outcomes (intended and non-intended) in different countries or sectors.[52]

Secondly, there are still relatively few actors in the community, and many of them are focused on open data as an "objective", rather than as a tool or a means for increasing impact or outcomes. Many local actors are mostly strong advocates of public data release and promoting the reuse of existing data. Few of them use open data in their work or engage in releasing their own data. Moreover, many local open data initiatives do not yet focus on traditional development sectors such as education, health and agriculture. Additionally, there is little connection between horizontal and vertical open data initiatives (e.g. national versus vertical sector-specific initiatives). Finally, there is still underinvestment in the sector in terms of both donor and State funding.[53] All this is true of the African open data community, and all this limits the contribution of open data to sustainable development in Africa.

b) Open data communities in Africa

The term open data communities in Africa refers to communities of actors (public, private, civil society, academics, citizen groups, etc.) engaging in work with open data initiatives, data portals and the mediated use of open data sets to ensure transparency and accountability on the part of governments, aid agencies and private-sector companies and for the general purpose of promoting sustainable development.[54] Like elsewhere in the world, they are still largely focused on open government data or the release of data produced by the public sector, such as official statistics. Nonetheless, open data communities in Africa also include actors focused on data from non-governmental actors, such as oil, mining and gas companies (open extractives), aid agencies (open aid), government procurement (open contracting) and scientific publications (open access). Thus, the actors within the African open data community include governments and governmental agencies, academic and research institutions, the private sector, civil society and non-governmental organizations, citizen groups, international development and aid agencies, especially the World Bank AfDB, and global open data communities.

48 Manyika and others, 2013.
49 World Bank, 2014.
50 Davies, Perini and Alonso, 2013.
51 Ibid. See also Boyera and Iglesias, 2014.
52 Boyera and Iglesias, 2014.
53 Ibid.
54 Ibid.; Davies, Perini and Alonso, 2013; Jesuit Hakimaki Centre, 2014.

A number of African countries have launched or are in the process of launching national open data initiatives or open government data portals, dedicated to making many, although by no means all, types of official statistics data open. They include Benin, Burkina Faso, Ethiopia, Ghana, Kenya, Morocco, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Swaziland, Tunisia, Uganda and the United Republic of Tanzania. Some communities provide data on governance and other non-traditional data, in addition to providing official statistics. For example, the open data initiative in Burkina Faso provides online information on company registration, legislation and election results, as well as the national map and transport timetables. Some municipal or local governments in Africa also have their own open data initiatives. Edo State in Nigeria has its own open data platform which provides more than 200 data sets from the State government and non-State actors. The city of Cape Town also has its own open data portal.

Kenya was the second African country (after Morocco) to launch a national open data initiative, in 2011. The Kenya Open Data Initiative is a national government programme whose goal is to make core government development information (demographic, statistical and expenditure data) available in a useful digital format for researchers, policymakers, ICT developers and the general public. It was launched with much dynamism and enthusiasm, catapulting the country into the forefront of the open data movement. According to the World Wide Web Foundation's 2016 Open Data Barometer, Kenya was ranked at position 42 out of 92 countries worldwide - the highest rank in sub-Saharan Africa. The Kenya open data portal has had more than 170 million page views since its launch in 2011. So far, 902 data sets have been published from 31 out of the 62 government agencies. It is the most visited government portal in the country.

In launching the Open Data Initiative, as a response to articles 34 and 35 of the Constitution of Kenya, which guarantee freedom of information and the right to access information, the Government was also seeking to reap social and economic benefits from open data. President Kibaki linked the constitutional imperatives with an economic rationale.

"I call upon Kenyans to make use of this government data portal to enhance accountability and improve governance in our country. Indeed, data is the foundation of improving governance and accountability. This is because reliable and timely data is the basis for determining whether the Government is delivering services effectively and accountably. Availability of data will, therefore, enable citizens to keep track of service delivery. This way the people can hold government service providers accountable for the use of public resources".

As in Kenya, the open government data movement in Africa has a national or sectoral focus, although as we have seen in Nigeria, there is a nascent municipal or subnational government dimension as well. This is because this movement is still largely driven by the commitment that data held or produced by government – data which in general tend to be national and sectoral in scope - should be open. In 2014, to supplement national open data initiatives in individual countries, AfDB launched the Africa Information Highway, which comprises two types of portals for each participating country: a statistical data portal and an open data portal. The statistical data portals contain official national statistical data that are disseminated and controlled by national statistical offices, while the open data portals contain data from national and other sources such as the United Nations and the World Bank and allow users to create and share content directly on the open data portal or through social networks. The AfDB initiative brands itself as a "one-stop centre" for development data in Africa.

To a large extent, the open government data movement in Africa reproduces much of the data within national statistical systems, although in a variety of countries they are sometimes repackaged or converted to use in novel ways or in localized circumstances on an isolated or ad hoc basis. For example, through World Bank and other donor initiatives, Africa-wide open data communities such as the African Media Initiative and Code for Africa have been implementing an open data literacy programme designed to enable mass media to translate open data into actionable intelligence for mass public consumption, and to inform decision-making at the national and local levels. Technically, Code for Africa is a citizen-based data community, a people-driven movement that aims to empower active citizenry

and strengthen civic watchdogs to help government shape and improve its services to citizens. It is driven and co-funded by grassroots citizen organizations and the mass media, and is focused primarily on building civic technology capacity within civil society and the watchdog media. This illustrates the overlap between some of the data communities, in this case between open and citizen-based data communities.

Non-State actors such as these are therefore critical components of open data communities in Africa. Others include private-sector, civil society and scientific data communities. The private sector plays multiple roles in open data, although many of them are still relatively small-scale or entail experimentation. There are private-sector companies that are committed to making some of their data open, especially in the extractive industry sector (e.g. the Extractive Industries Transparency Initiative and the Nigerian Oil Spill Monitor provided by the Nigerian Oil Spill Detection and Response Agency). There are those, especially technology companies, that provide platforms or software to support open data movements (e.g. IBM's Smarter Cities initiative and Google's traffic alerts in Kenya). Finally, some private-sector companies design and run open data initiatives. For example, in Mali, the ICT company Yeleman has launched an open data initiative that provides access to data sets on elections, urban infrastructure and primary schools in Bamako.[55]

Many civil society organizations in Africa are not only advocates for open data, but also users of open data. Some of them are not specifically focused on open data but have as their core mandates promotion of freedom of information, anti-corruption, transparency, accountability and good governance. They include Budgit in Nigeria, which uses an array of tools to simplify the budget and matters of public finance for the people of Nigeria in order to enhance transparency, accountability and participation in public finance. It has developed a tool – Tracka - which allows Nigerians in 17 States to post pictures of development projects in their communities. Budgit's project officers aid citizens without Internet access to communicate with their elected representatives.[56] Other civil society actors such as Map Kibera in Kenya and Ramani Tandale in the United Republic of Tanzania are specifically focused on open data and often integrate geospatial, crowdsourced and official data of relevance to particular communities, regions and peoples.

Within the scientific community in Africa, a number of open science platforms are springing up. An open science platform is an integrated set of arrangements that provides a policy, capacity-building and infrastructural framework for enhanced accessibility and impact. Such initiatives also focus on the creation of national open science forums. The African Open Science Platform, launched in December 2016 at the South African Science Forum, is an Africa-wide initiative which aims to promote the development and coordination of data policies, data training and data infrastructure. It seeks to enhance the value and exploit the potential of open data for science for the benefit of society. The platform also acts as a conduit for links with international open data and open science programmes and standards. The pilot phase is being supported by the South African Department of Science and Technology, funded by the National Research Foundation of South Africa, directed by the Committee on Data for Science and Technology of the International Council for Science, and implemented by the Academy of Science of South Africa.

As mentioned above, however, like the open data community globally, the African open data communities are relatively young. Their social, economic and political impacts have yet to be rigorously determined on a large scale. Many of them are still focused on open data as an objective rather than as a tool for enhancing social, economic and other outcomes. Few of them use open data in their own work or engage in making their own data open.

While many open data communities in Africa do focus on traditional development sectors and have great relevance to the Sustainable Development Goals (e.g. Map Kibera in relation to Goals 1 and 7), there is little connection and coordination between national and vertical sector-specific initiatives. Moreover, the lack of universal access to the Internet, differences in Internet bandwidth, differences in the use of technical standards (e.g. PDF

55 Alais, 2015.
56 http://yourbudgit.com.

files versus structured Excel spreadsheets versus machine-readable "linked data") and legal questions about copyright and "copyleft"[57] standards mean that different kinds of data tend to be "more open" with regard to one dimension or the other or with respect to different segments of the population. To the extent that open data generate any social, economic and political benefits, those who are able to access and use them in Africa are potentially more advantaged than those who are not. In this case, open data may have the potential to reinforce urban-rural, poor-rich and other socioeconomic divides. That is, open data are not by themselves inclusionary or even pro-poor. Proactive efforts must be made to make them so. Finally, most open data community initiatives in Africa are still donor-funded. This is not a sustainable situation.

2.2.6 Big data communities

There are multiple definitions and types of big data. The McKinsey Global Institute defines big data as "data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse".[58] The most famous definition of big data comes from Doug Laney, who defined it in terms of the volume (or amount) of data, their variety (or the different data formats such as figures, words, pictures, video, audio call logs, mobile banking transactions, online user-generated content, satellite images, financial or social transactions) and velocity (or the speed of data generation and processing).[59] The United Nations Global Pulse initiative describes big data as "a massive volume of both structured and unstructured data that is so large that it's difficult to process with traditional database and software techniques".[60]

From these definitions, a number of characteristics of big data emerge. Firstly, big data are characterized by the three Vs – large volumes and varieties of data and the velocity at which the data are generated and/or processed. The recent growth in the velocity, volume and variety of data comes from exponential increases in telecommunication bandwidth that connects a network of centralized and decentralized data storage systems, which are increasingly processed at great speeds thanks to digital computational capacities.[61] Secondly, big data come from everyone (governments, private-sector companies, civil society organizations, citizen groups and academic or scientific communities) and everywhere (social media sites, mobile phone GPS signals, sensors gathering climate information, digital pictures and videos posted online). Thirdly, big data merge different sources (data fusion) and are often accessible in real time.[62] Finally, the concept of big data goes beyond the increasing quantity and quality of data to include a focus on the analysis of the data for intelligent decision-making (big data analytics).

Applications of big data have been held to be potentially transformative in social, economic and business development and in various other areas. While this potential has been demonstrated to a large extent in the developed world, the full transformative potential of big data has yet to be realized in the developing world. In the developed world, uses of big data have included: [63]

57 "Copyleft is a type of licence that attempts to ensure that the public retains the freedom to use, modify, extend and redistribute a creative work and all derivative works (i.e. works based on or derived from it) rather than to restrict such freedoms. This is accomplished by the copyright holder granting irrevocable permission to the public to copy and redistribute the work in the same or modified form, but with the conditions that all such redistributions make the work available in a form that facilitates further modification and use the same licence." See Linux Information Project, Copyleft definition. Available at: http://www.linfo.org/copyleft.html. Accessed on 28 December 2016.

58 Manyika and others, 2011.

59 Laney, 2001.

60 United Nations Global Pulse, 2012.

61 Hilbert, 2016.

62 Real-time data are data that cover or are relevant to a relatively short and recent period of time, such as the average price of a commodity over a few days rather than a few weeks, and are made available within a time frame that allows action to be taken that may affect the conditions reflected in the data. Maaroof, 2015.

63 Brynjolfsson, Hitt and Kim, 2011; Lenard and Rubin, 2013.

Increasing firms' output and productivity (one study found that that United States firms that have adopted big data analytics have output and productivity levels 5–6 per cent higher than their other investments and information technology use would lead analysts to expect)
Predicting the spread of flu more rapidly than established medical research organizations such as the United States Centers for Disease Control, through the use of Google Flu, an online-search-based big data analytics platform)
Tracking health risks by combining airline records, disease reports and demographic data (as done by the Centers for Disease Control)
Studying whether mobile phone use increases the risk of cancer. For example, the Danish Cancer Society combined Denmark's national registry of cancer patients with mobile phone subscriber data to study whether mobile phone use increased the risk of cancer
Using data on retail transactions from hundreds of online retail websites to produce alternative price indices that are made available in real time, before the official consumer price indexes are published (the Billion Prices Project)
Using Google search engine data to provide accurate measures of unemployment and consumer confidence
Predicting housing market trends
Using billions of flight price records to predict the movement of air fares and offer customers low air ticket prices (Farecast)
Offering protection against risks of credit card fraud or terrorism (Palantir).

While the scale and impact of efforts to harness big data in Africa have not been as large as in the developed world, the age of big data for development has nevertheless arrived in the developing world, including Africa. As a recent study by professional services company Accenture put it: "Big data has arrived at the base of the pyramid - the largest but poorest social-economic group which currently comprises 4 billion people globally".[64] The United Nations Global Pulse initiative argues that if properly harnessed, big data can support global development in three main ways:

- Early warning systems, for example in the case of disease, famine or humanitarian crises and disasters
- Real-time awareness through "a fine-grained and current representation of reality which can inform the design and targeting of programmes and policies"
- Real-time feedback, or the ability to monitor a population in real time and to understand how policies, programmes and projects are performing.[65]

The big data community in Africa is still relatively small, but it is growing rapidly. In 2014, an IBM report on information technology pacesetters focusing on Kenya and Nigeria found that 40 per cent of businesses were in the planning stages of a big data project, compared with a global average of 51 per cent. The study also found that 12 per cent of Kenyan and Nigerian firms already had live big data projects, compared with the global mean of 13 per cent. Moreover, the big data community in Africa is not limited to the private sector. It also includes government, academia, civil society and other actors. The IBM Smarter Cities project is an example of the practical application of big data analytics to urban and infrastructure development in Africa. A good illustration of this project is the Outthink Urban Planning Initiative in Nairobi, Kenya (box 5).

Nature of the big data community in Africa

The big data community is, albeit still relatively small, growing rapidly in Africa. Its size can be attributed to a number of factors. First and foremost is the state of technological infrastructure, human and financial resources. Martin Hilbert argues that the prerequisite for making big data analytics work for development is "a solid technological (hardware) infrastructure, generic (software) services and human capacities and skills. These horizontal layers are the requirement sine qua non … They can be unequally distributed, leading to a development divide. Once available, the horizontal layers can be employed to analyse different aspects and kinds of data, such as words, locations, nature's elements and human behaviour, among others".[66] To the extent that African countries lag in these, they are more likely to lag in applications of big data for development.

64 Long and Brindley, 2013.
65 United Nations Global Pulse, 2012.
66 Hilbert, 2016.

Box 5: Outthink Urban Planning Initiative in Nairobi, Kenya

Nairobi has a population of 5 million people and urbanization is outpacing infrastructure development. The Outthink Urban Planning Initiative was conceptualized following a discussion between Nairobi County Environment Executive Evans Ondieki and an IBM Research Lab team led by Ms. Aisha Walcott-Bryant. The discussion focused on how the city could best manage its fleet of dust carts. At the time, the vehicles were performing only one trip a day, and were unable to collect most of the rubbish in the city. The major contributors to this problem were traffic congestion, poor road conditions and the inability of supervisors to track the dust carts and to devise different strategies aimed at greater efficiency.

IBM designed a scalable solution that seeks to generate data for use in urban infrastructure planning within developing countries in order to address some of the most immediate needs, such as traffic congestion and road safety. Outthink was a customized solution that leveraged cheap and easily accessible technologies - specifically mobile phones, which were combined with sensors to build a tracking device which was then installed on dust carts.

The sensor suite is able to collect data on the speed of the vehicle, altitude, road quality, location of potholes and speed bumps along road networks, in addition to change in terrain from paved to dirt roads.

Prior to the Smarter City initiative, Nairobi city was collecting only 800 tons of rubbish a day. This figure has almost doubled since the introduction of the fleet management system, and the city is now able to collect more than 1,400 tons in a day. The city has also used the data to identify road blockages, accidents, detours, unmarked speed bumps and hazardous potholes, enabling quicker repair. This has assisted the city engineers and planners to design better solutions which help in reducing the inefficiencies in the current infrastructure.

Indeed, experimentation and innovation in big data in Africa are concentrated in some of the countries with the deepest Internet and mobile penetration. Apart from Senegal, many of them are English-speaking countries of eastern and southern Africa, together with Ghana and Nigeria in West Africa. This is the second feature of the big data community in Africa. Some countries, such as Ghana, Kenya, Senegal and South Africa, tend to be engaged in relatively large or diverse experimentation and innovation with big data (and hence have relatively large big data communities), while other countries have thinner big data communities and limited experimentation with big data.

Thirdly, the size of the big data community in Africa can also be attributed to socioeconomic, demographic and political contexts. Different social, political and economic systems may give rise to significantly different rates of diffusion of big data between industrialized and developing countries, among developing countries and within developing countries.[67] In other words, different social, political, and economic systems can have differential impacts on the diffusion and effective utilization of big data in key development areas (for example, differences in laws, legislation and policies governing the production, ownership, dissemination and use of data). It is perhaps not surprising that some of the most active and dynamic big data (as well as open data) communities in Africa are found in countries that were quick to reorient their national development policies and legal frameworks towards harnessing data for development, such as Kenya, Rwanda and South Africa.

Fourthly, the big data community in Africa is still characterized by limited interactions within the community itself. Much of the experimentation under way typically involves two or more actors and partners within countries. Generally, there is still limited systematic or structured interaction among the big data communities within countries or between countries. Part of this has to do with the difficulty in accessing data, especially among government agencies and private-sector actors, for example because of privacy concerns, the personal and commercial sensitivity of data and other legal and proprietary restrictions. Part of it also has to do with the different types of big data and differences in technological, human and financial capacities among various stakeholders.

There is greater interaction between the big data and open data communities in Africa, as well as between the big data and citizen-based data communities. This is partly because the open and citizen data communities have few if any restrictions on the use, reuse or redistribution of data. Most analysts do believe, however, that the large-scale development and practical application of big data in Africa will happen only if private-sector and official statistics data communities are willing to share vastly larger quantities of data than

67 Ibid. See also Kshetri, 2014.

has been the case so far. These two communities are not only two of the largest holders and producers of data in Africa, they are also among the largest consumers of data on the continent. Even with the open data and open government data movement in Africa, the national statistical system in many African countries has still not made most of the data that it holds open. The same is true of the private-sector data community.[68]

Opportunities and challenges

The above examples, coupled with other global experiences, suggest that applications of big data could be transformative for the achievement of sustainable development in Africa, especially in the areas of poverty, agriculture, health, energy, education, innovation and infrastructure, and climate change. Big data can help unlock and generate new insights from the vast quantities of data that are already held by governments, as well as data held by citizen groups and other stakeholders. Data fusion can enable big data to fill in key gaps within official statistics in Africa, helping to expand the conception of who and what counts and is counted. The early warning, real-time awareness and feedback qualities of big data can help improve programme, policy and project implementation.

While much of the experimentation and innovation in big data in Africa is of great relevance and potential to all the 17 Sustainable Development Goals (as some of the above examples suggest), a number of challenges to the effective application of big data to sustainable development remain. Some of these relate to technological infrastructure and human and financial resources, as already mentioned. Others relate to social, economic and political contexts. For a number of African countries, in order to benefit from the full transformative power of big data for their development, they will need to invest in and/or somehow develop or acquire significant technological capabilities and technical expertise. African governments and stakeholders will need to make explicit and deliberate choices to harness big data for development and to create enabling legal, legislative, policy and other environments.

Other challenges relate to the scale, scope and quality of big data in Africa. The big data community in Africa is still relatively small, working on relatively micro, isolated or pilot projects and on a relatively limited range of issues. If big data are to be effectively applied in the pursuit of sustainable development in Africa, this community will have to be able to cover most African countries and regions, and focus not just on relatively small-scale and isolated projects but on national, sectoral and subnational issues as well.

Perhaps more importantly, the quality of big data analytics will need to rival its relevance and scope. Even after all these conditions are met, it should be noted that big data cannot and will not replace the official statistics system. Rather, they will complement official statistical data. The big data and official statistics data communities will need to work more closely. As the United Nations Global Pulse initiative has argued, big data do not and should not be expected to contain all the answers to human problems. They also have their own limitations and biases, which must be understood and taken into account. Finally, "real-time information does not replace the quantitative statistical evidence which governments traditionally use for decision-making, but if understood correctly, it can inform where further targeted investigation is necessary, or even inform immediate response if necessary, and thus change outcomes like nothing else can".[69]

2.2.7 Citizen-based data communities

Around the world, citizens and citizen-based groups and organizations have been involved in collecting, processing, disseminating and using different types of data - social, economic and scientific – for some time. However, in the last two decades there has been exponential growth in these citizen-based data communities. This has been aided by the combination of advances in ICT, greater democratization, growing transparency and accountability movements, and increasing global efforts to harness different types of data for a whole range of different causes. Most citizen-based data community initiatives seek to achieve greater public accountability and transparency, social, economic and environmental change and enhanced scientific knowledge and public understanding of science.

Depending on whether they are collecting, processing, disseminating or using data, and whether they

68 Kshetri, 2014. See also United Nations Global Pulse, 2012 and Maaroof, 2015.
69 United Nations Global Pulse, 2012.

engage in socioeconomic or natural science activities, citizen-based data groups have been categorized variously as "citizen science", "crowd science" and "citizen data-generating" communities. Generally speaking, citizen groups working with data in the natural sciences, especially conservation biology and astronomy, have a much longer organized tradition than most citizen groups focused on socioeconomic and political data. For example, citizen-based natural history data investigations date back to the seventeenth century.[70] For this reason, the terms "citizen science" and "crowd science" have come to refer to citizen-based data collection, processing and dissemination in the natural sciences, while "citizen-generated data" is commonly used in citizen-based data generation in the socioeconomic and political worlds.

Citizen science refers to public participation in scientific research, with particular reference to cases where members of the public partner with professional scientists to collectively gather, submit or analyse large quantities of data. Crowd science refers to the crowdsourcing of citizen-science-based data.[71] Citizen-generated data are typically defined as data that people or their organizations produce to directly monitor, demand or drive change on issues that affect them. They can be produced through crowdsourcing mechanisms or citizen reporting initiatives, with or without organization by civil society groups.[72] Citizen-based data communities are therefore diverse, and can include and overlap with scientific, civil society and open data communities.

As with the other data communities, the potential applications of citizen-generated data can be transformative. Globally, the ability of citizen-based data communities to collect, process and distribute scientifically, socially, economically and politically useful data has been demonstrated. One study has identified 388 unique citizen-science-based biodiversity projects in which between 1.36 and 2.28 million people volunteer each year, making estimated annual in-kind contributions of between $667 million and $2.5 billion.[73] These projects have yielded a total of 446 scientific publications in high-profile journals such as *Bioscience*, *Conservation Biology*, *Trends in Ecology and Evolution* and *Proceedings of the National Academy of Sciences*.

Since 2007, a project called Galaxy Zoo has been enlisting the public in classifying images of space captured by the Hubble space telescope. Over 150,000 people classified more than 50 million images in the first year alone. Professional scientists would not have been able to accomplish this feat even with their sophisticated computer algorithms.[74] In Nepal, a citizen data group known as Local Interventions Group has collected, cleaned up and used data to track how public and donor money was spent on the humanitarian response to the Nepal earthquake of 2015. The Group visualized this data into a story which was told to communities previously left out of the decision-making process, helping them to engage more actively in demanding accountability and transparency in the implementation of the earthquake response initiatives.[75] In Argentina, at least two citizen-based data groups (Yo Quiero Saber and Policrowd) have experimented with generating and disseminating political data on public office holders, candidates for political office and top civil servants, in order to help enhance transparency and accountability in political governance and public service.[76]

These examples suggest that citizen-based data groups are an integral part of the data revolution. Citizen-generated data in areas such as climate change, environment, biodiversity and land use could be potentially transformative in the pursuit of the Sustainable Development Goals. Citizen-generated data are seen as having the potential to contribute to achieving sustainable development by:

- Helping identify and filling in key gaps in official statistics
- Reducing the costs of data collection, processing and dissemination
- Fostering data integration, fusion and harmonization

70 Miller-Rushing, Primack and Bonney, 2012.
71 Bonney and others, 2016.
72 Civicus, 2017; Adieno, 2016.
73 Theobald and others, 2015.
74 Bonney and others, 2016.
75 CEI, 2016.
76 Ibid.

- Enhancing the ownership and relevance of data and improving governance
- Contributing to scientific knowledge, greater public understanding of science and greater science-society relations.

In various parts of the world, there is evidence that citizen-based data can achieve many of these aims, at least on a small scale. One study has highlighted the achievements of citizen science as follows: "Scientific outcomes of citizen science are well documented, particularly for data collection and data processing projects. … We find limited but growing evidence that citizen science projects achieve participant gains in knowledge about science knowledge and process, increase public awareness of the diversity of scientific research, and provide deeper meaning to participants' hobbies. We also find some evidence that citizen science can contribute positively to social well-being by influencing the questions that are being addressed and by giving people a voice in local environmental decision-making".[77]

In spite of this potential, citizen-generated data and citizen-based data groups continue to face a number of challenges. Many of these relate to the accuracy, reliability, consistency and quality of citizen-generated data. Questions about the rigour of the methods employed in relation to citizen-generated data abound. These include questions about sampling, in particular criticisms of the crowdsourcing method. Is it scientifically representative, rigorous and objective? There are also questions about the scale, replicability and sustainability of citizen-generated data. Many citizen-based data efforts tend to be small-scale, short-term, isolated or one-off events.[78]

However, criticisms of citizen-generated data extend beyond disagreements over methods and quality to include questions about the quality and accuracy of the claims that proponents of citizen-generated data often make. Does citizen science empower communities and individuals to improve their well-being? Does citizen science engage underserved or marginalized segments of the population? Does it engender greater trust and ownership of data? Does it lead to the democratization of science and decision-making? Most of these questions remain unanswered. If citizen-generated data are to be fully harnessed for sustainable development, citizen and other data communities will need to provide answers to many of these questions.

Citizen-based data communities in African countries are similar to their counterparts elsewhere in the world. They are diverse, growing, experimenting with filling in the gaps in official statistics and scientific data, and continue to face questions about the accuracy, reliability, quality, scale and sustainability of their data. Citizen-based data groups in Africa overlap with open data communities (e.g. Map Kibera and Ushahidi in Kenya), scientific data communities (e.g. citizen science and crowd science initiatives in environmental and natural resources investigations in Kenya, South Africa and Uganda) and civil society data communities (e.g. Civicus-based and DataShift-supported citizen-generated data initiatives in Kenya and the United Republic of Tanzania, and the African Media Initiative and other African citizen journalism projects). And they are in the early stages of exploring potential collaboration with official statistics communities (e.g. Twaweza in both Kenya and the United Republic of Tanzania).

Citizen-based data groups in Africa collect, process and distribute data on social, economic, political, environmental, scientific and humanitarian issues. Ushahidi, for example, best described as a well-known Kenya-based crowdsourcing platform that gives voice to marginalized communities in times of crisis such as natural disasters or violence emanating from conflicts in an effort to uphold justice, transparency and human rights, aggregates reports, creates visual maps and enables real-time or near-real-time response to crises by non-governmental organizations, government agencies, journalists, emergency responders and other stakeholders.

Ushahidi (Swahili for testimony or witness) was formed in response to the violence that erupted soon after the December 2007 Kenyan Presidential election results were announced. Approximately 1,400 lives were lost and between 180,000 and 600,000 people were displaced from their homes in just 59 days. In an effort to alleviate the crisis, several Kenyan software developers and bloggers joined hands in building the first version of the platform. According to an anal-

77 Bonney and others, 2016.
78 Ibid.

ysis carried out at the time by Harvard University's Kennedy School of Government, the platform was found to be more effective than mainstream media, as it was able to report on fatal and non-fatal violence in addition to being able to collate information from both rural and urban rural areas. Verification of the information from eyewitnesses, which was sent either by email or in text messages, was achieved through triangulation with additional sources from Ushahidi's partners including international media, government sources, NGOs and Kenyan journalists and bloggers.[79]

Ushahidi is an open-source platform and is licensed under the GNU Lesser General Public License. It went global shortly after its Kenyan release. In 2008, for instance, it was used to track anti-immigrant violence in South Africa. It was also used in 2009 to track pharmacy stock-outs in several African countries including Malawi, Uganda and Zambia. During the same period, Al Jazeera television used it to collect eyewitness reports during the war in Gaza between Palestinians and Israel. It has also been used to manage earthquakes in Chile and Haiti in 2010, and to track corruption-related cases in partnership with organizations such as Transparency International in Macedonia in 2011. It was even used to assist in providing real-time alerts during the 2010 winter storms in Washington, D.C. and wildfires in Russia.[80] The Ushahidi platform has had over 90,000 deployments worldwide, with a global outreach of more than 20 million people. It has been used to monitor elections, generate alerts during earthquakes and improve humanitarian aid responses. It has received numerous awards, including the 2012 Global Adaptation Index Prize, the 2013 MacArthur Award and the 2016 Classy Award for Social Innovation.

In eastern Africa Trac FM radio also generates real-time citizen data to identify socioeconomic and political trends. In collaboration with NGO and broadcasting partners, it collects, processes and disseminates real-time citizen-generated data on socioeconomic and political matters. Trac FM works with local radio stations and communities to identify their most pressing concerns. It has software which allows radio hosts and presenters to put questions to their audience during live shows. Members of the audience can share their answers free of charge by sending text messages to the Trac FM system. Responses are automatically categorized and visualized in graphics and maps. Radio hosts can then discuss poll results in real time with their guests and audience. The full data set is later aggregated in a range of comprehensive graphs and maps for further analysis by the radio station and by Trac FM's NGO or government partners.[81]

Since 2011, Trac FM has worked with 13 partners in, Kenya, Somalia, Uganda and the United Republic of Tanzania. It has received over 1 million text message responses from more than 394,000 users. Some of Trac FM's projects include:

- Polling to find out why more men than women in Uganda start businesses
- Documenting Ugandans' views on the police
- Determining the main reason why girls drop out of primary school in Uganda
- Conducting a survey of what Ugandans think should be done about refugees
- Asking Ugandans whether sex education should be taught in schools
- Investigating the shortage of contraceptives in Uganda's public health-care system
- Enquiring into poor security, unemployment, foreign aid and refugees in Somalia.

Citizen-based data communities are also contributing to environmental and natural resource monitoring and management in eastern and southern Africa. In Kenya, South Africa and Uganda, volunteers monitor and record data on habitats and ecosystems such as forests, mangroves, lakes and rivers. They also undertake the counting and mapping of various animals including sea turtles, bats, owls, frogs, lizards, butterflies, birds, zebras, elephants, lions, leopards and cheetahs. One such citizen-based data group in South Africa, Cape Citizen Science, asks nature lovers to become "pathogen hunters" while on nature walks by recording dying fynbos plants and collecting samples of dead plant material. African national parks also engage in stimulating citizen science and citizen-based data collection. Snapshot Serengeti asks volunteers for help in sorting photographs gathered by hundreds of camera traps in Serengeti National Park in the United

79 Mungai, 2016.
80 www.Ushahidi.com.
81 Communication Initiative Network, 2016.

Republic of Tanzania, while South Africa's Kruger National Park occasionally asks visitors to help monitor endangered species like wild dogs and cheetahs.

The work of the Animal Demography Unit at the University of Cape Town is indicative of the overlap and interactions between some citizen-based and scientific data communities in Africa. The Unit feeds data from citizen scientists into atlases and distribution maps of birds, dung beetles, frogs, scorpions, spiders, butterflies, dragonflies, lacewing moths, sea stars, mushrooms and orchids. It has an innovative virtual museum which contains over 6,000 photographic records provided by hundreds of citizen scientists. The Extreme Citizen Science initiative in University College London, another excellent example of growing collaboration between citizen science and conventional scientific data communities in Africa, links various scholars in the university with citizen scientists in the Democratic Republic of the Congo.[82]

Perhaps the most ambitious and dynamic citizen-based data organization in the field of socioeconomic and political data in Africa is Twaweza, an East African citizen-based data organization. Twaweza's mission is based on the premise that "citizens of East Africa can bring about change themselves, rather than waiting for governments, politicians, donors or NGOs to do it for them". Accordingly, Twaweza seeks to provide practical information to everyone in order to foster high-quality independent media and citizen monitoring services. In Kenya, Uganda and the United Republic of Tanzania, the Twaweza-supported Uwezo programme runs Africa's largest annual citizen assessment to monitor children's learning levels across hundreds of thousands of households. Through its Sauti za Wananchi initiative, Twaweza also runs Africa's first nationally representative mobile phone survey.[83]

Citizen-generated data initiatives can harness the power of collective intelligence to contribute in areas where key data are missing. This is what the Uwezo initiative has done in East Africa with respect to primary education data and the Millennium Development Goals.

These examples demonstrate both the opportunities posed by, and the limitations of, citizen-generated data in Africa. These are similar to the strengths and weaknesses of citizen-generated data elsewhere. The opportunities include: the potential to fill in gaps in official scientific statistics and other types of data; the potential for transformative innovation, especially in data fusion and promotion of open data communities; unique capabilities for real-time data and feedback that can be quite critical in decision-making and policy, programme and project implementation, as well as in responses to humanitarian crises; and enhancement of public participation and ownership of data, and the cost-effectiveness of data collection, processing and dissemination. In terms of the Sustainable Development Goals, based on the above examples, there is potential room for citizen-generated data and citizen-based data communities to contribute to Goals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14 and 15.

The challenges are centred on quality, accuracy, reliability, sustainability, sampling and scale.

While there is some interaction between citizen-based and scientific data communities in Africa, interaction between citizen and official statistics data communities in most African countries is still largely non-existent. There is also limited interaction among the various citizen-based data communities themselves. Until this begins to change, the potential applications of citizen-generated data to the Sustainable Development Goals effort will remain minimal.

The distribution, level of activism and relative acceptance of citizen-based data communities in Africa are determined by different social, economic, political, demographic and technological contexts. As with open data communities, there appears to be an overconcentration of citizen-based data groups in Anglophone eastern and southern Africa, together with Ghana and Nigeria. To the extent that these data communities have a relatively low presence or profile in many African countries, this limits their potential contribution to the pursuit of the Sustainable Development Goals.

82 See https://uclexcites.wordpress.com/about/ and http://vmus.adu.org.za/.
83 See http://twaweza.org/go/what-is-twaweza.

However, even after many of these challenges have been addressed, it should be understood that the purpose of citizen-generated data is not to replace or supplant official statistics. Like big data, a key value addition of citizen-generated data lies in their ability to complement and fill in the gaps in data from official statistics, scientific and other communities. The complementarities can come in such areas as speed (i.e. real-time diagnosis and feedback), scope (i.e. reaching difficult-to-reach, marginalized or often excluded segments of the population), cost-effectiveness (i.e. in data dissemination) and engendering trust and ownership within national data ecosystems.

Box 6. Uwezo case study: Citizen-generated data supporting improvements in free primary education in Kenya

Prepared by DataShift

Millennium Development Goal 2 sought to "achieve universal primary education".

What was the education challenge in Kenya?
Between 1980 and 2000, Kenya experienced declining enrolment at all levels of education, especially at primary school level. About 53 per cent of children who had enrolled in Standard 1 (ages 4-5) in 1994 did not complete the primary school cycle.

How did the Millennium Development Goals help to resolve the education enrolment challenge?
Following adoption of the Goals in 2000, and in a bid to meet the requirements of Goal 2, the Government of Kenya developed a policy of free primary education, which was formally introduced for all schoolgoing children in January 2003.

Under this policy, levies and fees were abolished for all primary school students, and the Government also made new funds available to purchase learning materials and for capacity-building for education managers. It earmarked 5.4 billion shillings (approximately $54 million) from the national budget to implement the policy, and raised an additional 4 billion shillings from external resources to support the programme. As a result, enrolment of school-age children rose from 5.8 million in December 2002 to 7.2 million in May 2003 - an incredible leap in a very short period of time.

Link between free primary education and the Millennium Development Goals?
Unfortunately the implementation of this policy was never widely understood (at least by the general public) as relating to the Government's fulfilment of its commitments under the Millennium Development Goals. Rather, it was seen as more of a politically driven "gift" by the then newly elected Government. This gap in understanding was in large part due to a lack of publicly accessible and/or insufficiently promoted information on the Goals, especially in terms of how they directly translate to real public benefits and impact the lives of ordinary citizens. The fact that so many citizens were unaware of these links, coupled with their incorrect perception of the politics surrounding free primary education, limited their collective ability to hold the Government to account for those commitments. They did not wish to appear unappreciative of the "gift".

What challenges emerged from free primary education?

- Schools experienced serious overcrowding as the Government made it compulsory for children of schoolgoing age to attend classes while not expanding the infrastructure to accommodate this increase
- The capitation fund (an amount allocated on a per-pupil basis to cater for primary education needs within primary schools) was not sufficient to serve the needs of the additional millions of children who entered the system
- The pupil-to-teacher ratio increased drastically, as the number of teachers could not be increased to keep pace with the increased student enrolment
- Despite success in increasing enrolment rates to such a high level, the quality of the education itself deteriorated, as indicated by a drop in both literacy and numeracy rates.

How did citizen-generated data help to highlight these education challenges in Kenya?
Uwezo, a five-year initiative that aims to improve literacy and numeracy among children aged 6-16 in Kenya, Uganda and the United Republic of Tanzania, uses an innovative approach to social change that is citizen-driven and accountable to the public. It has made use of citizen-generated data produced from annual citizen surveys to demonstrate that despite impressive increases in student enrolment rates, the overall quality of education in Kenya and East Africa has in fact deteriorated.

Uwezo's 2014 report *Are Our Children Learning? Literacy and Numeracy Across East Africa, 2013* highlighted these deficient outcomes across the three countries, while annual reports since then have consistently proved that pupils were not learning the core skills expected for their age and grade.

- More than two of every three pupils enrolled at Standard 3 level in East Africa fail to pass the Basic English, Kiswahili or numeracy tests set at Standard 2 level
- Improvements in basic literacy and numeracy as children progress through the education system occur only slowly, implying that the quality of learning remains low throughout primary school.

Uwezo's study further revealed that by the time East African children reached the last year of primary school, one out of five had still not acquired basic literacy and numeracy skills (although some did end up acquiring some skills through the curriculum followed in their later school years). In Kenya, 6 out of 10 children aged 10-16 possessed literacy and numeracy skills at Grade (Primary) 2 level, compared with 5 out of 10 in the United Republic of Tanzania and 4 out of 10 in Uganda.

Initially there was significant resistance to the Uwezo reports by Kenya's Ministry of Education, which responded by commissioning its own independent verification process. When its own results proved shockingly similar to Uwezo's findings, the Government finally started paying attention and giving credence to the initiative's learning assessments.

How can citizen-generated data support the implementation of the Sustainable Development Goals, as well as monitoring progress towards them?
Uwezo's work is primarily centred around Goal 4, which focuses on education. More specifically, it connects the targets of educational quality and learning outcomes, indicators for which many developing countries still have inadequate data. Its approach also has potential to bolster the efforts to measure Target 8.6, by supplementing official data on unemployment and school enrolment rates, as well as to assess the quality of training that young people in East Africa are receiving. Furthermore, Uwezo is contributing to the improvement of local governance by engaging civil society in monitoring the delivery of public services.

About DataShift
DataShift is an initiative started by Civicus that builds the capacity and confidence of civil society organizations to produce and use citizen-generated data.

Box 7. Building a solid foundation for a sustained data revolution in Africa: Country data compacts

Prepared by the African Population and Health Research Center

The paucity of accurate, reliable and timely data has been a major constraint to the effective targeting of international, national and donor investment. It also impacts monitoring and evaluation of development programmes within and across countries in Africa. Country data compacts represent one approach for enhancing a country's capacity to produce accurate, timely and relevant data.

Country data compacts

A compact is a formal, multi-year agreement with a clear national road map, roles and responsibilities for the government and its partners to co-fund specific, measurable, achievable, realistic and time-bound activities. These contextual frameworks for governments and donors could be used to fund progress on core national statistics, engage civil society and the private sector in new ways, and mobilize new technologies for data collection and dissemination. Country data compacts could go a long way towards improving development data quality and tackle entrenched political economy challenges.

Compacts are a way to add value to what countries have already been doing to improve their national statistical systems through the Strategy for the Harmonization of Statistics in Africa or the road map outlined in the Africa Data Consensus. A compact is a natural next step to actualize the road map in a country, tailored to the local context. It outlines roles and responsibilities for government and its partners, including co-funding, and offers a way to grow beyond existing efforts such as the national strategies for the development of statistics: not to replace them, but as a complement to them and a way of addressing challenges encountered by existing efforts. A compact would foster engagement from both producers' and users' perspectives across data communities, and, it is hoped, include modalities of moving beyond statistics to better utilize data.

Despite existing efforts and leadership to improve data and statistical systems in Africa, political will has not yet been enough to reach a tipping point, especially on prioritizing national investment in statistical systems. Political will and leadership as well as financing are essential for action on development, data and statistics, and other critical issues facing countries. Country champions - including strong leadership in the shape of directors general of national statistical offices and statisticians general - have to fight for the prioritization of country investment as a top national budgetary priority.

Delivering a data revolution means forging a new relationship of accountability and cooperation among governments, donors, producers and users of statistics, and including other data communities beyond the traditional producers of official statistics. Whether it is through use of a country data compact or other means, African countries must accelerate action to address the core problems facing national statistical systems that underlie poor-quality data in the continent. The time is now.

Achieving a country data compact: Steps and activities

1. Establishing and formalizing a country data leadership team

As with other initiatives including national strategies for the development of statistics, securing high-level political and technical national governmental support is the first, critical step of this process. A country data compact "champion" is essential to secure high-level commitment to the process and to broker trust. Team leadership is formalized by establishing a country data leadership team, which is charged with coordinating the development and implementation of the compact. The team will include senior officials of the national statistical office, those from other relevant agencies and staff who commit to dedicate a predetermined amount of time to the effort over the course of the implementation of the compact. The national statistical office will also serve as the secretariat for the team in order to support its rules of governance, a work plan, vision and strategy, and a communication strategy.

2. Conducting an objective inventory of existing national capacity

Countries must conduct an objective institutional assessment of a country's data production and dissemination systems to document how the national statistical system can be improved. This assessment will be used to establish a country baseline against which progress will be measured. It may also be used to identify and prioritize activities designed to improve national statistical capacity rapidly. Previous assessments carried out under national strategies and the PARIS21 Advanced Data Planning Tool* will provide a good starting point.

3. Mapping the data ecosystem and convening key data actors

A mapping of the data ecosystem will be conducted ahead of a National Data Forum. The forum, hosted by the national government through the national statistical office, will convene national experts and leaders from government, donors, academic and research centres, civil society organizations and the private sector to secure commitments to contribute intellectually, financially and technically towards improving national data systems. The forum will also draw upon the expertise of institutions engaged on the data revolution at regional and continental levels in Africa, including the African Union Commission, AfDB, the World Bank and ECA, all of which are already working to improve statistical

* http://www.paris21.org/ADAPT.

production in Africa. Their involvement will be crucial to the likelihood of success of the compact in identifying best practices, avoiding the repetition of past missteps, but also providing a boilerplate compact that can be tailored for use in other countries in which these same institutions are working to improve data.

4. Building consensus on the goals and objectives of a compact and developing a work plan

Compacts may target one or more specific, concrete objectives, such as achieving the autonomy of the national statistical system, generating sustained domestic funding, coordinating data generation and use across sectors, improving data management and validation systems, improving open data and data-sharing policies and platforms, or a combination of such objectives, grounded in results from the country self-assessment and country priorities identified from consultation.

Soon after the forum, a series of professionally facilitated meetings will assist the country data leadership team to formalize the country data compact elements and work plan on the basis of forum inputs and resource commitments; the plan will include timelines, costing and metrics for tracking progress against an established baseline. Once a work plan has been developed, vetted and adopted, the country will, through the National Data Forum, identify specific resources to carry out defined activities in accordance with a predetermined timeline.

5. Monitoring and evaluation

As a part of its work plan and funding request, the country data leadership team should plan for monitoring and evaluation against the baseline established as a part of step 2 to measure progress over time and allow for course correction if needed. Communications and dissemination activities will be included as part of the plan, with deliberate efforts being made to document lessons learnt from the country compact to inform action by other countries considering the same. Annual briefings could be developed by the leadership team (e.g. State of the Data Revolution) to provide updates.

6. Sustaining data improvements

Long-term financing is the big outstanding question, and this will rely almost entirely on generating political will for sustained domestic financing. One way of sustaining the capacity and prioritization gained under a compact and associated funding is to embed it in each country's long-term development plans or visions (such as Vision 2030 in Kenya, Vision 2020 in Rwanda, etc.).

Investment in data systems is for the long term, not a one-off. Collaboration between donors will lead to faster, better-coordinated efforts. Donor and domestic investment with short-term funding horizons of even five years is likely to be insufficient to meet long-term needs. Capacity-building is a process: its strategy needs to be targeted, and should start to move from the supply side to the demand side.

About the Center

The African Population and Health Research Center is a leading pan-African research institution headquartered in Nairobi, Kenya, that conducts policy-relevant research on population, health, education, aging, urbanization and related development issues across Africa. It engages policymakers and other stakeholders to help ensure decision-making across the continent is informed by research. The Center is also committed to developing the next generation of globally competitive African scholars through numerous research-capacity-strengthening initiatives and partnerships with top African universities and research institutions.

3. Data innovation in Africa: Selected case studies

The foregoing discussion shows that national data ecosystems in Africa are characterized by considerable innovation and experimentation. These are taking place across all data communities. Much of the experimentation and innovation entails efforts aimed at promoting the quality, inclusion, ownership, relevance, timeliness, reliability, openness and use of data. This chapter presents highlights of such innovations across different data communities in selected countries. Many of these case studies involve collaboration between two or more different data communities. They speak to what is possible within the context of more dynamic, functional and collaborative national data ecosystems in Africa.

3.1 Rwanda: Innovations within the official statistics system[84]

From Community Score Cards to Citizen Report Cards to the Rwanda Governance Scorecard

The national statistical system in Rwanda has undergone considerable innovation and experimentation over the last 10 years, including the launch of Citizen Report Cards and Community Score Cards, and more recently the Rwanda Governance Scorecard, all intended to improve the quality of social, economic and political governance and to enhance transparency and accountability within the public service. Specifically, the Citizen Report Cards and Community Score Cards are aimed at improving efficiency, policy effectiveness, accountability and participation in decentralized levels of administration. They are designed to obtain feedback on public service delivery at both macro and micro unit levels. They are also intended to improve dialogue between front-line service providers and the users or recipients of public services. In this way, they can be seen as tools for increasing public participation and empowerment in decision-making at local levels.

Citizen Report Cards are the tools of participatory surveys that provide feedback on user perceptions of the quality, adequacy and efficiency of public services. Community Score Cards are qualitative monitoring tools that are used for local monitoring and performance evaluation of services. Both have been in existence in Rwanda on a pilot basis since 2004. In 2010, the Government introduced the Rwanda Governance Scorecard, which measures governance performance against eight indicators: rule of law, political rights and civil liberties, participation and inclusiveness, safety and security, human and social development, control of corruption and transparen-

84 This section draws heavily from Nzayisenga, 2016.

cy and accountability, quality of service delivery and economic corporate governance.

The Rwanda Governance Scorecard is customized to Rwanda's dynamic governance needs and technical and technological capabilities. It is also benchmarked against global governance research methods and standards. It draws from new sources of local data such as expert perceptions from Citizen Report Cards and Community Score Cards, together with more conventional statistics and surveys. It thus constitutes an interesting experiment in data fusion or the merging of different types of data in decision-making. The Community Score Cards, Citizen Report Cards and Rwanda Governance Scorecards have all been developed in partnership between the official statistics community and other data communities.

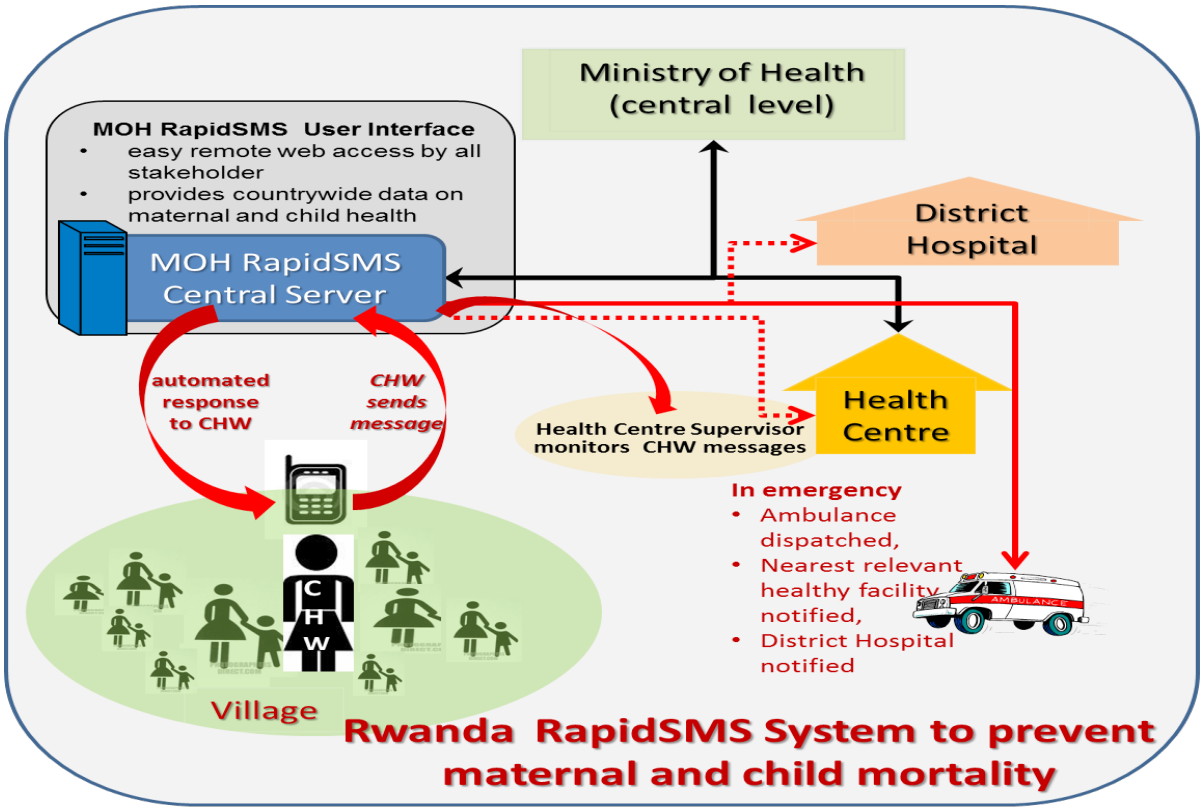
RapidSMS and maternal and child health

RapidSMS is a platform for data-gathering and group communication using the short message service (SMS) on mobile phones. The aim is to tackle the problem of slow data transmission in the field of maternal and child health among community health workers in Rwanda. It was developed in 2009 by the Ministry of Health in partnership with the United Nations Children's Fund (UNICEF).

How it works

The community health worker uses his or her phone to register in the system. Prior to the introduction of RapidSMS, only hard copies (see photograph below) were completed and used to track pregnant mothers. It took time for community health workers to monitor and refer clients, and feedback was received only after a delay. Now, using SMS mes-

Figure 4. Information flow in the RapidSMS system



sages through mobile phones, the message and feedback occur almost in real time.

The RapidSMS system helps community health workers to track pregnant women and to monitor antenatal care by identifying women at risk and referring them for medical attention. It also helps to identify and address the reasons behind the deaths of women and young children at the community level.

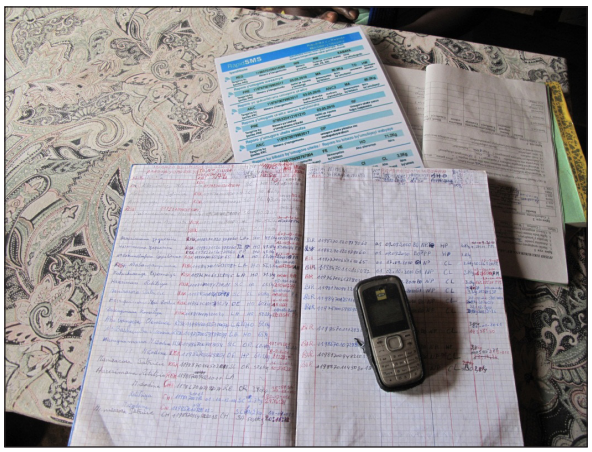
Impacts

This innovation has improved maternal and child health in Rwanda. Community health workers are able to report in real time on pregnant women at risk, antenatal care visits and deaths of mothers or children. With a simple message sent by phone, an ambulance can be made available in an emergency even in the remotest parts of Rwanda. Benefits include:

- Increased antenatal visits and care
- Increased identification of children with malnutrition and other challenges related to maternal and child health
- A lower number of risky home deliveries
- Between 2010 and 2015, the infant mortality and under-five mortality rates fell from 50 per cent to 32 per cent and from 76 per cent to 50 per cent respectively.

3.2 United Republic of Tanzania: Innovations in technology, systems and tools in the national statistics system[85]

Section 17 of the United Republic of Tanzania's Statistics Act (2015) gives the National Bureau of Statistics the mandate to coordinate the national statistical system. The Bureau sets standards for collecting, analysing and publishing statistics to ensure uniformity in quality, adequacy of coverage and reliability of statistical information. It is also responsible for issuing a code of practice for official statistics, including professional standards to be followed by all agencies authorized to produce official statistics. As a custodian institution, the Bureau is also responsible for maintaining a comprehensive national databank by drawing from sectoral databanks developed by other public agencies.



The Government has invested in upgrading its data technologies and platforms. For example, within the national statistical system the following data platforms and tools can be found:

- *Socioeconomic database*. The database captures official statistics from various government ministries and changes in various socioeconomic variables including income, consumption expenditure, literacy level, asset ownership, access to education and health services. Depending on data sources, it allows disaggregation of data at the national, district and subdistrict level. The database also provides web dashboard for quick analysis on key indicators.
- *Disability statistics database*. Captures information or statistics on persons with disabilities, including employment, education and household income.
- *National Data Archive*. This database provides data for development. It includes the agricultural census, basic statistics and information on the construction industry, the Business Register Survey and the Employment and Earnings Surveys.
- *Food and agricultural statistics*. This is a statistical framework and applied information system for statistical data and metadata on food and agriculture originating from different sources.
- *Local Government Monitoring Database*: A system designed to enable local and regional au-

85 This section draws heavily from Mdadila, 2016.

thorities and the Ministry responsible to monitor development initiatives at the local level.
- *Sectoral/thematic management information systems*: These exist in sectors such as water (e.g. water point mapping), education and health.

The National Bureau of Statistics disseminates much of the country's data, with exception of data produced by other government units.[86] The dissemination of official statistics is guided by the following principles:

- Equal access to official statistics among all users (including the requirement to announce the release date in advance)
- Official statistics must be relevant for users and be user-friendly
- Official statistics must be accurate, reliable, consistent and transparent.

The Bureau still mainly performs data collection using traditional methods such as manual collection through questionnaires, although there is a movement towards digitizing data collection and storage. For example, the recent Demographic and Health Survey used tablets as a way to improve efficiency, consistency and timely dissemination of the findings. Furthermore, it is expected that the upcoming Household Budget Survey will use mobile digital devices in the field in order to improve the quality and timeliness of data collection. Increasingly, the Bureau is acquiring the necessary ICT infrastructure for data collection and management, but capacity gaps remain.

3.3 Kenya: Citizen-generated data: Ma3Route

Ma3route provides data on traffic patterns and accidents in Kenya, with the aim of improving transit in the major cities. It has more than 500,000 daily users who access the platform's news feeds either through native mobile apps, the Web, Twitter or Facebook. Ma3route uses crowdsourced data generated by citizens while in transit. Ma3route was created in response to a report by the National Transport and Safety Authority which found that pedestrians accounted for the overwhelming majority of traffic-accident-related deaths in Kenya. Policymakers did not know the nature and location of these accidents. They often relied on accident data from the police, hospitals, insurance companies or the media. However, these sources do not provide complete data, since not all accidents come to their attention. Ma3route was designed, at least in part, to help address this problem.

Ma3route, in partnership with the Transport and Safety Authority, the National Road Safety

Trust and Zusha, a project supported by Development Innovation Ventures at the United States Agency for International Development, started an initiative which would help generate data on accident locations in real time. They invited citizens to use the platform to help create a map of accidents in Nairobi, by using the platform to post accident information including time, location, description and photo. Ma3route then developed an algorithm to validate the correctness of the information. It checked whether the accident had been the subject of more than one report, and whether there was a photo of the incident. At the end of the first six months, 7,817 accidents had been reported, 1,324 of which had been geo-tagged and verified. The data suggested that 42.5 per cent of the accidents had happened within 500 metres of a pedestrian bridge, which indicates that the bridges are underutilized. The accidents also involved more public service vehicles than private cars.

Following these findings, in an effort to obtain additional insight, Ma3route made an effort to harness data from the Kenya Red Cross and National Transport and Safety Authority . The Authority's data are available on the Kenya open data portal, making it easy for Ma3route to access. Government agencies also have access to Ma3Route data, which they can use for policymaking and enforcement, in an effort to reduce road carnage. For instance, the Government can create incentives to enhance the use of footbridges by pedestrians, by introducing fines or deploying police officers to accident hotspots.

86 United Republic of Tanzania, 2010.

Box 8. Automatic identification and counting of elements to monitor progress towards the Sustainable Development Goals

Prepared by the United Nations Global Pulse Lab Kampala

The 2030 Agenda for Sustainable Development challenges us to think and work differently to achieve sustainable development. An open mind is needed to look at the data revolution as an opportunity to tap into new sources and technologies that complement current approaches for analysis and monitoring of the Sustainable Development Goals. Governments, partners and stakeholders must work together in Africa to ensure that technical capacities are built to explore the use of innovative technologies and new types of data. Innovation and partnership to create collaborations to tap into new types of data are key to exploring the opportunities that these can offer.

Over the past six years, at Pulse Lab Kampala, the third Lab of the United Nations Global Pulse network, we have worked with partners in the public and the private sectors to help governments to tap into the data revolution. Global Pulse catalyses work across a data ecosystem: creating frameworks for the United Nations to make better use of new sources of data, demonstrating how they can be used, and showing how such data can enable the United Nations and governments to deliver its mandate.

The use of digital devices whenever we go about our daily lives is increasing at an unprecedented pace in Africa. You can buy and sell goods with your phone, send parcels, map unrest; there are social media, mobile banking, and the promise of cashless societies and digitized land records all across the continent.

There has been an explosion of mobile phone use in Africa. A decade ago, Nigeria had around 100,000 phone lines; today there are around 140 million active mobile phone lines. The explosion of mobile phone ownership is also leading to exponential growth in Internet use on the continent. Research indicates that in Africa, mobile data use will increase 20 times by the end of the decade, twice as much as the anticipated global expansion, with estimates that there are around 340 million Internet users in Africa with a penetration rate of 30 per cent.*

As early as 1999 Uganda became the first country on the continent where the number of mobile subscribers passed the number of fixed-line users. The Communications Commission estimates that for every landline there are 18 mobile phone connections in the country. It is estimated that 12 million people had access to the Internet in 2016, accounting for around 30 per cent of the population.** Facebook users alone totalled 150 million across the continent, with some 2.2 million in Uganda.

The world, and Africa with it, are in the midst of a data revolution. The amount of data produced globally and regionally is more than doubling every year. As people communicate through mobile devices or social media or by searching on the Internet, they produce a vast amount of digital data known as big data. In addition, various new technologies and sensors allow for the exploration of innovative and exciting new digital data sources, such as satellite imagery, radio content and telecoms data. What is reported, but also what people say, what people do and what sensors measure can provide us with interesting big data that have the potential to save lives, inform policies and enable authorities to respond faster and more effectively to situations on the ground. Data derived from mobile phone devices can be critical in analysing the spread of a disease outbreak. Radio content analysis can give us exceptional insights on such topics as health service delivery, and even contribute to reporting on local disasters.

As a result of the increased use of digital devices in our daily lives, a wide range of new data sources are available to support official statistics in the continent. Making the deluge of new data sources and the methodologies derived from them useful for development requires a new approach to official statistics. The United Nations Statistical Commission agreed in March 2014 to create a Global Working Group on Big Data for Official Statistics. At its first meeting, the Group acknowledged that using big data for official statistics is an obligation for the statistical community, based on the fundamental principle of meeting the expectation of society for enhanced products and improved and more efficient ways of working. So how do we change mindsets and apply big data for official statistics and to advance the Sustainable Development Goals? The key is to look at the potential of new data sources to complement traditional statistics with real-time, targeted updates. While we know that big data cannot represent universes of analysis in the way that other data do, we can still approach big data with a mind open to finding the opportunities and advantages that they offer. At Pulse Lab Kampala, we aim to illustrate how big data can be used to generate statistics to monitor progress towards the Sustainable Development Goals and complement official statistics. And to do so we are developing with partners a range of digital applications.

A prototype for automatic counting

While significant progress has been made on reaching Sustainable Development Goal 1 and eradicating poverty in Uganda, a large percentage of the population (43 per cent) remains poor and highly vulnerable to poverty.*** Uganda has experienced growth and diversification of the non-agricultural economy in the last few years. Yet a vast majority of households in rural areas are vulnerable to crop losses, livestock diseases, illnesses or price fluctuations. With low

* Internet World Stats - Usage and Population Statistics. Available from http://www.internetworldstats.com. Accessed on 23 February 2017.

** Ibid.

resilience to face risks, the household economies are volatile. Addressing the vulnerability of poor households to poverty requires understanding of risk factors and resilience levels. Equally important is monitoring of the fluctuations at the household economy and poverty levels. Household surveys are currently the main source of data in this regard, but they take place at five-year intervals and sample less than 0.1 per cent of the population.

A proxy indicator of poverty identified by the Uganda Bureau of Statistics is roofing material. Traditional thatched roofs harbour pests and diseases and require a high degree of maintenance. As the household economy improves, families often upgrade their dwelling by changing from the traditional grass thatch to iron sheets. We can observe changes in the roofing materials of households and in the landscape if we observe satellite imagery over time. Manual detection and counting of structures observed with satellite imagery is a common practice used, for example, to support emergency operations like the displacement of refugees.

As the Data Scientist at Pulse Lab Kampala John Quinn explains, "If the human eye can identify a change, trained software can also do it". Pulse Lab Kampala has developed a prototype to measure progress towards Sustainable Development Goal 1 – End poverty in all its forms everywhere - using "automated roof counting" through satellite imagery. The prototype uses image processing software that identifies and counts roofs and the type of material they are made of automatically. With this prototype, we aim to transform a manual process into an automatic process to generate statistics from high-resolution satellite imagery data captured with remote sensing. Like other applications developed with partners at Global Pulse, this digital application is based on artificial intelligence and machine learning.

The prototype has been developed with partners at the University of Edinburgh in the United Kingdom. Further development of the prototype includes the following steps:

- Selecting appropriate high-resolution satellite imagery for the country within a time frame that will show changes in the type of roofing materials for a specific area
- Collecting ground truth information, where available, to verify the accuracy of the software, and manually tagging it for evaluation areas where it is not
- Training of deep learning models to distinguish different types of roofs, and evaluation of performance.

An operational prototype could generate statistics in real time and in an automatic manner from satellite imagery on an ongoing basis. Remote sensing data are collected continuously and globally, using a consistent process in all countries and regions of the world. Hence, technology to derive statistics from such data can naturally be scaled up to other countries.

The statistical community is conscious of the fact that before introducing big data into official statistics, and in order to take advantage of these innovative data sources, including their application to monitoring and reporting on the Sustainable Development Goals, it needs to adequately address issues pertaining to methodology, quality, technology, data access, legislation, privacy, management and finance, and provide adequate cost-benefit analyses. Pulse Lab Kampala and its partners aim to address this issue through the implementation of the pilot project to test the prototype.

The digital application could help national statistical offices to fill data gaps between surveys, to potentially reduce the frequency of surveys, to reduce the costs of data collection, to reduce the burden on respondents and to provide timely data at a more disaggregated level for informed decision-making.

The Lab aims to test the prototype in 2017 with partners in government to enhance real-time monitoring and support the implementation of policies and programmes to achieve Sustainable Development Goal 1, as follows:

- Measurement of poverty. The innovative methodology will complement existing statistical tools that use surveys and primary data collection to assess poverty levels
- Monitoring of poverty and vulnerability to poverty. Assessments and surveys based on non-digital data provide in-depth analysis to support decision-making. This "static" picture of reality can be updated with a regular flow of timely information and data
- Analysis of real-time data on changes in the landscape can provide information and offer new insights on trends, shocks or coping strategies of communities.

About United Nations Global Pulse

Global Pulse is a flagship innovation initiative of the United Nations Secretary-General on big data. Its vision is a future in which big data are harnessed safely and responsibly as a public good. Its mission is to accelerate discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action. The initiative was established based on a recognition that digital data offer the opportunity to gain a better understanding of changes in human well-being, and to get real-time feedback on how well policy responses are working. The first

*** Uganda, 2012.

innovation lab in Africa, Pulse Lab Kampala brings together data scientists, data engineers, partnership specialists, academics and technical experts to generate high-impact data analysis tools to address development challenges. These innovative tools support United Nations partners and government in anticipating and responding to poverty, impacts of natural disasters, epidemics and food security by leveraging new sources of digital "big data" (such as social media, mobile data and online information) and real-time analytics. Through partnerships, and in alliance with the Government, Pulse Lab Kampala acts as an ecosystem catalyst.

3.4 South Africa: The Victims of Crime Survey: An example of a national statistical office going beyond its traditional domain

Crime statistics are both emotionally and politically highly charged and subject to intense scrutiny, debate and criticism. Given South Africa's history and the fact that South Africa has one of the highest relative crime levels in the world, its crime statistics often attract close scrutiny. Official crime statistics are released by the South African police services. However, the systematic underreporting of certain serious crimes (such as rape) and "less serious" crimes (such as theft and robbery), coupled with questions about internal reporting procedures, have made the official crime statistics a subject of intense questioning.

Statistics South Africa has demonstrated a willingness to engage with statistics beyond its traditional domain, especially where strong social and societal development urgencies are manifest. Thus its Victims of Crimes Survey is of particular interest. After Statistics South Africa conducted a first such survey in 1998, the Institute for Security Studies carried out two more in 2003 and 2007. In 2011, Statistics South Africa resumed data collection, and adopted a continuous data collection methodology from 2013.

The Survey is notable in acknowledging that there is a wide gulf between actual statistics about crime and how crime is experienced or perceived, by both its victims and community members. Thus, more than half of the 90-page report is devoted to household perceptions of crime and safety in general, and victim and law enforcement support services, including the services offered by the police, the courts and correctional institutions. The Survey also has dedicated sections on hard-to-measure crimes such as trafficking or corruption. It also analyses reasons why some crimes go unreported.

One of the key findings of the latest report (2014-2015) concerns the gap between the perception and the reality of crime in South Africa. In general, the public perceives the crime rate to be higher than reported by official statistics. The report also provides stunning insights into the underreporting of crime. Property-related crimes (such as housebreaking/burglary) are reported in only 52 per cent of cases; theft of personal property in only 34 per cent; livestock 32 per cent. Crime impacts South Africans in a variety of ways: 15 per cent of households feel unsafe walking within their neighbourhoods during the daytime. This number increases to a whopping 69 per cent at night. As many as 37 per cent of South Africans will not go into open spaces or parks unaccompanied, and more than half of households (52 per cent) have physical protection for their houses.

Focused statistical surveys like these are a useful and insightful addition to the raw crime statistics typically published by the police or justice departments. The willingness to engage with these types of data attests to the political courage and commitment to using statistics for societal improvement and development by South Africa's national statistical system. Statistics South Africa is taking this forward by launching a new series of reporting on gender and vulnerable group statistics.

Box 9. Gender equality and the African data revolution

Prepared by the World Wide Web Foundation

Sustainable Development Goal 5 aims to achieve gender equality and empower all women and girls. A crucial target under this goal is to enhance the use of ICT to promote this empowerment. According to UN Women, only 13 per cent of countries dedicate a regular budget for gender statistics. Data2x, an initiative of the United Nations Foundation to improve gender data, has identified data gaps in 28 areas - from girls' learning outcomes to sexual violence in conflicts to unpaid work and more. We cannot work towards gender equality without improving our data, and by extension knowledge and understanding of these issues and how they affect women and girls. This lack of data is particularly pronounced across the Global South, and is a priority concern for organizations working on global inequality. Without these critical data sets, it will be hard to measure the true size and nature of the social and economic challenges for women and girls, and difficult to construct policies that identify and exploit opportunities for them.

Corinne Maurice of Maurice Communications Marketing wears her health data as a necklace while demonstrating Pass Santé Mousso+ to TechMousso Jury. Photo by Charly Kodjo

Web Foundation research shows that poor urban women in the developing world are nearly 50 per cent less likely to access the Web, and therefore unlikely to access online data and information that empowers them. We must close the gap in access to both the Internet and high-quality data if we want to change course to achieve Sustainable Development Goal 5.

TechMousso

Between March and July 2016, the World Wide Web Foundation, in partnership with the Millennium Challenge Corporation and Data2X, worked on a gender data initiative - TechMousso. The name was drawn from "technology" and "mousso", the word for "women" in Bambara, a language spoken across West Africa.

TechMousso aimed to:

- Generate excitement concerning the use of open data, gender data, and uses and tools for data
- Bring together the technology community, civil society and government actors to facilitate partnerships to make better use of gender data
- Build understanding of the challenges around gender and development, as well as the definition and sources of gender data
- Provide resources and networking opportunities for stakeholders
- Generate gender data for use in informing policy in Côte d'Ivoire
- Assist Côte d'Ivoire in its work on development of women and girls and achievement of the Sustainable Development Goals
- Increase the effectiveness of and empower community organizations working with women and girls in Côte d'Ivoire
- Increase demand for and use of gender data.

The initiative was implemented in three phases:

- Phase 1, Workshop for civil society organizations: In a one-day workshop, organizations working with women and girls, government ministries and other data stakeholders came together to identify specific needs that could be helped by better use of gender data.
- Phase 2, Competition: This phase included a call for solutions and a media campaign, and resulted in entries being received from over 80 teams. A longlist of 40 teams were then invited to pitch to a panel of judges, followed by an official launch where 20 teams were selected to work further on their solutions during a boot camp. Each of the 20 teams received an Internet connectivity kit as well as support from experts in statistics, data privacy, data visualization and mapping from government and civil society. Unlike in other initiatives, teams were given a month to work on their projects, to allow time for deep thinking and development.
- Phase 3, Awards gala: Entries (e.g. apps, visualizations, algorithms or data sets) were submitted for judging. Awards were presented to the top 10 teams at a gala event.

Following the kick-off workshop, TechMousso distilled stakeholder data needs to five key themes:

- Women's economic activity
- Women's health

- Women's legal and political capacity
- Human capital
- Education.

Teams were required to build a solution on one of these themes. The first prize was won by Mafubo for a solution that allows networking and data-sharing among maternity centres, hospitals and health districts to manage the availability of spaces and medical equipment to monitor pregnant women, among other things. The second prize went to Women Health+ Pass, for a connected item of jewellery that carries patient data and helps medical personnel take better and faster care of citizens. Dblamou, a statistical system for gathering data on gender-based violence, was placed third.

Blogs about the project are available on the World Wide Web Foundation website (webfoundation.org), and a short video documentary is also available (www.youtube.com/watch?v=soHdlRizcCY).

Lessons learned from TechMousso

Partnerships ensure that solutions are well-rounded and engage all parties needed to make change

TechMousso taught us that gender data initiatives at the national level require dynamic, functional and trusted partnerships. The first partnership is with the ministry in charge. This is the basis for most other partnerships. The national statistical office is also a key partner in any given data initiative in a country. By default, it is the official data holder for the country and also the facilitator of data-gathering as and when gender data collection moves forward. The head of the agency needs to be consulted and partnered with ahead of implementation. These parts of government are critical to ensuring that initiatives gain traction and can be translated into possible policy reforms.

Officials from the Access to Information Commission, the National Institute of Statistics, the Women's Compendium, the United Nations Educational, Scientific and Cultural Organization and the World Wide Web Foundation (seated) with a cross section of civil society gender actors from across the country during the first ever gender data national round table. Photo by Charly Kodjo

In Côte d'Ivoire, the Compendium of Women's Expertise was an early partner with TechMousso. Throughout the life of the project, the Compendium network served as a great clearing house for information on the state of play across the country. Partnership with a network of civil society actors in the subject area is key. For gender data, the Compendium was an ideal partner. Though it is not technology-oriented, its insight into the data needs of its over 70,000 network members proved highly useful for solution providers.

Often building a data culture and data capacity is the biggest win

Despite government and ministerial engagement with TechMousso and access to information in the country, there was still no guarantee that participants could obtain the information they needed. In many cases, government officials are not sure exactly which data sets fall into the public domain. Officials at different levels are cautious about operating within the bureaucracy. Even for issues that are clearly within their mandate, for which on-the-spot decisions can be taken, they will still prefer the longer way: a formal letter, a request for authorization, a face-to-face meeting, referral to the hierarchy, a follow-up meeting - to be sure they have covered everything and are able to release the information.

Data collection among rural women is still carried out using pen and paper. Shea butter producers queue up for a TechMousso session in Doropo, near Bouna, Côte d'Ivoire. Photo by Charly Kodjo

Beyond access is the issue of data quality. Often, data sets are incomplete, unavailable or just non-existent. According to the World Wide Web Foundation's Open Data Barometer - which not only measures the openness of data but also evaluates the quality of data available in the countries - Africa is lagging far behind the rest of the world in having high-quality data, let alone making such data publicly available.* The national statistical office alone cannot fulfil all the data needs of a country. No matter how willing its authorities may be, their operational framework does not allow for flexible data production to keep pace with data needs in an environment of rapid social and technological changes. Data demands may be met by other data holders, like the telecommunications industry, chambers of agriculture, mines or commerce, and even cooperatives and associations.

* http://opendatabarometer.org/3rdedition/regional-report/africa/. Accessed on 1 November 2016.

Nation-wide inclusive engagement will yield most benefits

TechMousso was implemented in Côte d'Ivoire, a country of 23 million people spread over 31 regions. Though most of its economic activities are clustered around the economic capital, it was important to engage civil society organizations and technology experts from across the country. National inclusive engagement was embedded in the fundamental design of the initiative, and was implemented throughout the phases. Of the 20 teams which competed for the Grand Prize, 4 were from outside the economic capital. This cross-country coverage of the initiative was highly appreciated by stakeholders and competitors alike.

Recommendations

- Experience with TechMousso has produced the following key recommendations. African governments which wish to ensure that high-quality open data yield benefits for all citizens and policymakers should:
- Review, assess and design survey methods that take into account gender dimensions
- Ensure that statistics or data departments in line ministries are operational and resourced
- Eliminate institutional barriers within government to share data across ministries and statistical offices, including at the subnational level
- Make a commitment that national statistical offices will have open data by default, as envisaged in the Africa Data Consensus, and provide gender-disaggregated data
- Adopt the Open Data Charter, following the lead of Sierra Leone, to provide a framework for open data and better data use in policies adopted in pursuit of the Sustainable Development Goals
- Construct and continually reinforce a multi-stakeholder data ecosystem built on trust, allowing for more dynamic engagement of all stakeholders in the data value chain
- Incentivize data contributors, communities and actors who otherwise will not engage in enriching the national data ecosystem, by providing the data that they produce or hold
- Include a gender section in countries' open data platforms
- Ensure that the indicators of Goal 5 are translated into national objectives, tracked and regularly reported upon.

About the World Wide Web Foundation

The World Wide Web Foundation was established in 2009 by Web inventor Sir Tim Berners-Lee to advance the open Web as a public good and a basic right. We're building a future in which the Web empowers everyone, everywhere, to take part in building a fairer world.

We want everyone to be able to access, understand and use data to change their society for the better, which is why open data, and in particular gender data, are such an important part of our work. We're working to spark an open data revolution where everyone is counted and included.

4. Conclusions and recommendations

4.1 Conclusions

Africa is undergoing a data revolution. This is the result of deliberate political, economic, social, technological and policy choices and actions on the part of multiple stakeholder groups, including, central and local government, the private sector, civil society, citizen groups and scientific or academic communities. Collectively (at the continental level) and individually (at the country level) there is high-level political will and commitment to harness data for accelerated sustainable development. There is also significant goodwill and commitment from a number of private-sector companies, NGOs, media entities, citizen groups and academic institutions.

Various data communities are engaged in considerable experimentation and innovation on how to harness the data revolution in ways that would transform social and economic development, climate change adaptation and mitigation, and environmental and natural resource management. At the heart of this experimentation and innovation has been a reconceptualization and redefinition of what and who officially counts, is counted, how, by whom and for whom. This question extends beyond a simple focus on statistical techniques and methods, or even the notion of statistical capability. It is fundamentally a question about national development priorities, and accordingly what matters and is worthy of counting, planning for and monitoring.

The United Nations Sustainable Development Goals, the African Union's Agenda 2063 and most national long-term development plans now conceive of development more broadly to include economic, social and environmental dimensions. The data revolution in Africa responds to this more expansive concept of development. In a comparative historical perspective, this represents a fundamental shift away from using data for narrowly economic purposes to using data as a tool for promoting inclusivity, participation, integration, transparency and accountability in all dimensions of development.

It is one thing to have a clear plan of national development and a vision of the data ecosystem necessary to realize it. It is another thing altogether to create the legal, technological, human and financial resource architecture and infrastructure to realize this aim. This is one of the current fundamental deficits or gaps within national data ecosystems in Africa. For the most part, the political will is there. The goodwill and commitment of non-State actors are there too, and a number of measures that are changing the data landscape in Africa have already occurred as a result of this twin dynamic. These include legal, legislative and policy reforms as well as technological innovations and other forms of experimentation. However, changes within national data ecosystems are not occurring as fast or on a large enough scale as the response to the development challenges demands.

A number of obstacles remain, including:

- Uneven data coverage, accessibility and use - by issues, regions and population segments. Significant segments of the population, and issues important to them, are still excluded or not fully captured by many national data ecosystems in Africa
- A mismatch between data demand and data supply. Issues or areas for which data are required do not often necessarily align with the types of data that are in supply. Data are still not available in a timely, accurate or relevant manner. Data are still mostly supply-driven and funded by external sources, leading to a situation where donor or funder priorities take precedence over national priorities
- Lack of disaggregated data and the storage of data in sector-specific silos. There is a lack of harmonization of data collected by different groups and in different formats
- Data are still not significantly open by default in most African countries, nor are data largely accessible in usable form to most stakeholder groups
- Technological, human and financial resource constraints.

There are many reasons for this. Legal, legislative and policy reforms take time, where they work as intended in the first place. Successful technological innovations require not just technological feasibility but social acceptability and organizational or commercial viability. Capacity to harness technological applications and to implement development policy is predicated on the existence of skilled human capital. All of this requires money. By one estimate, globally, collecting data for all the 169 targets of the Sustainable Development Goals would cost $254 billion, nearly twice total annual expenditure on foreign aid.[87] In Africa, an unpublished report produced by the heads of African national statistical offices in 2015 estimated that producing harmonized statistics to track progress towards the 17 Sustainable Development Goals would cost a minimum of $1.54 for every one of Africa's 1.2 billion people. This figure would rise to $4.33 per person with efforts to obtain more reliable numbers, and would increase further to $8.11 per person by 2030, when the population is estimated to reach 1.6 billion. Of course not all decision-making by governments requires official statistics. However, Africa's statistical requirements also go far beyond what is needed for the Sustainable Development Goals, so these figures are likely to be conservative.

Africa must still overcome significant challenges if it is to benefit from the full transformative potential of the data revolution. Data communities in Africa tend to be more vibrant in countries with enabling institutional, technological, political and economic environments. These include legal, legislative and policy frameworks that anticipate and provide for data production, processing and dissemination by various data communities. Such frameworks also include the principles of inclusivity, high quality, relevance, timeliness, openness and accessibility to data production, processing, ownership, dissemination and use. The Africa Data Consensus embodies these and other principles and actions that would facilitate the creation of enabling environments for harnessing of the data revolution. A number of African countries have already reformed or revised their legal, legislative and policy regimes. Nevertheless, a lot of work remains to be done, especially with respect to governing frameworks for new forms, types or platforms for data such as open data, big data and citizen-generated data, in addition to creating conducive environments for potential integration and interoperability among these various data systems and platforms.

Not only must these reforms continue and be consolidated, but the various data communities need to seize these new legal, legislative and policy opportunities to work in collaboration with national statistical offices and other public data agencies to advance the vision of desired national data ecosystems in their countries. While governments, through national statistical offices, will continue to play a leading and guiding role within national data ecosystems, the desired national data ecosystems are not likely to be achieved if the other data communities do not play proactive roles in the effort to realize them. There is enormous scope for collaborative and cooperative arrangements between private-sector and public-sector entities as well as among public, private, civil society, citizen group and academic

87 Jerven, 2014.

data communities. Public-private partnerships such as those between Orange and the Governments of Côte d'Ivoire and Senegal, and between IBM and the city of Nairobi, might help in enhancing the technological, human and financial resource architecture and infrastructure within data ecosystems in Africa.

African governments and other stakeholders will need to make significant investment in data technologies, tools and platforms, and in human and financial resources. Virtually all African data communities and systems remain underfunded and understaffed and employ either suboptimal or outdated data technologies and platforms. The functional autonomy of the national statistical system, which is the objective of many policy and legal reforms across Africa, cannot simply be brought about by legal or policy diktat. It must also be backed up by skilled expertise, requisite technologies, sufficient funding and the ability of the national statistical offices to set their own budgets and work plans.

Efforts are under way in most African countries to align national development plans with the Sustainable Development Goals. Explicit efforts to this effect have commenced, for example, in Côte d'Ivoire, Kenya, Nigeria, Senegal, South Africa and the United Republic of Tanzania. As a result of investment in efforts to produce data to monitor progress towards the Millennium Development Goals, a number of national statistical systems in Africa have some capacity to produce data that can be used to capture the progress of what might be called the social and economic development elements of the Sustainable Development Goals (i.e. Goals 1, 2, 3, 4, 6, 7, 8, 9 and 10). With the exception of South Africa, which has the strongest national statistical capacity on the continent, in general, data production, dissemination and use with respect to Goals 11, 12, 13, 14 and 15 remain relatively weak in many countries. In the United Republic of Tanzania, for example, most of the indicators in these areas are new to the national statistical system. A total of 27 indicators in these areas will require new data collection systems, with 11 of them needing new systems with respect to Goal 15 alone. The cost of data collection for the Sustainable Development Goals and national development priorities will remain very high if it is to be met by the official statistics community alone. This is where collaboration among different data communities and the coordination, harmonization and integration of data from different sources and communities might be helpful in not only reducing costs but filling in key gaps in official statistics.

If the other data communities are to contribute to the harnessing of the data revolution at scale, they will need significant investment in data technologies, platforms and tools and in human and in human resources. Thus far, much of the data produced by non-State data actors have been relatively small-scale, isolated and irregular. The production of large-scale, regular, sustainable and high-quality data will require significant investment by these communities. Skilled and sufficient human resources are critical to high-quality, timely and relevant data.

The rise and growth of non-State-based data communities will not replace or diminish the role and functions of national statistical offices and the national statistical system. Rather, these communities complement and fill in the gaps within the national statistical system. In many cases, these communities produce different types of data and focus on issues, areas or regions that are typically either difficult to capture with conventional statistics or are generally excluded by official statistics.

How the official statistics community relates to other data communities is going to be critical to the degree to which Africa harnesses the full transformative potential of the data revolution. Will the official statistics community sustain its commitment to making official statistics open by default? Will it be willing to sanction high-quality, relevant, reliable, accurate, accessible and timely data produced outside the national statistical system as part of official statistics? Ultimately, will the national statistical system embrace a shift in its role as the creator and overseer of a new statistical landscape or ecosystem in which the field of data production and dissemination is opened up to both State and non-State actors? How these questions are ultimately resolved in individual countries will determine the degree to which they harness the full transformative potential of the data revolution.

While the relationship between the official statistics community and other data communities is going to be central, the interaction within and between these data communities themselves is also going to be

critical. There is currently limited interaction within and between different data communities. The result is unnecessary duplication, overlaps and an inability to develop standards, protocols and best practices that would lead to increased harmonization and integration. Greater interaction within and between data communities has the potential to reduce the costs of data collection, processing and dissemination.

In spite of its dominance within the African national data ecosystem, the national statistical system does not capture data on much of the informal economy in Africa and the people who engage in it, especially the poor and the marginalized. This is one of the gaps that civil society and citizen-generated data fill in many African data ecosystems, although even they do not claim to capture data on all segments of the population. But they do reach parts of the informal economy and population segments that the national statistical system usually finds difficult to reach.

In spite of the growth of civil society and citizen-generated data, in many African countries there remains some reluctance to use such data, especially on the part of the national statistical system, owing sometimes, but not always, to questions about their quality, accuracy and consistency. There is an enormous opportunity to strengthen the quality, accuracy and consistency of this type of data. A key advantage of this type of data, especially in the humanitarian sector, has been its timeliness and ease of accessibility. Civil society and citizen-generated data form a growing part of the national data ecosystem in Africa. However, the degree of dynamism of this community varies between countries depending on the depth of Internet and mobile phone penetration, political systems and literacy levels. In countries such as Ghana, Kenya, Nigeria, Rwanda, Senegal and the United Republic of Tanzania, this is an increasingly dynamic and innovative sector of national data ecosystems. Civil society actors tend to produce their own data but also to use and disseminate data generated by others including government, academia, the private sector and citizens. In this sense, they play a bigger role as brokers of different types of data in Africa, across almost all the 17 Sustainable Development Goals. Citizen-generated and crowdsourced data in Africa are relatively new but growing fast. So far, they seem to be concentrated in the humanitarian and governance sectors, although in a few countries (e.g. Kenya, Nigeria, Rwanda, South Africa and the United Republic of Tanzania) they have extended to health, education, water and other public service sectors.

National data ecosystems in Africa stand to benefit from the greater sharing, availability or accessibility of private-sector data. The challenge remains how to make these often proprietary data available for social or public good in a way that does not diminish their commercial viability for their producers. African countries need to seek ways and models for unlocking much of the data held by the private sector within the continent.

4.2 Recommendations

Against this background, this report makes the following recommendations:

- The importance of proof of premise. The need to conduct assessments on the full range of social, economic and environmental impacts of the data revolution in Africa in order to provide an evidence-informed basis for some of the significant investment and reforms that are going to be needed within national data ecosystems in Africa.
- The need to complete legal, legislative and policy reforms in ways that anticipate and allow for the harnessing of the data revolution for accelerated sustainable development in Africa.
- The need for significant investment in data technologies, platforms and tools (including general investment in Internet, mobile and other digital technologies) as well as human and financial resources in data and development, including investment in data science, computing, STEM, statistics, social sciences, economics, law, etc.
- The importance of greater collaboration and coordination within national data ecosystems in order to reduce the costs of data collection and help to enhance data quality and use. As non-State actors continue to play increasing roles in data collection and processing, and as the quality, relevance and timeliness of data produced outside official statistics improves, the need for greater collaboration among different data communities and greater coordination within national data ecosystems will become stronger.

The onus is on national statistical offices to provide this leadership.

- The need to enhance the functional autonomy of the national statistical system. In order to enhance and preserve its professionalism and integrity and the quality of its work, the national statistical system requires functional autonomy and a lack of politicization of its work. This entails financial independence and the ability to establish its own work plans and budgets. Also required are legislation or legal provisions that protect its staff from political interference.
- Non-State-based data communities should act proactively and responsibly to contribute to the creation of functional, vibrant and coordinated national data ecosystems. The responsibility to seize the expansive policy, legal and legislative space now lies with non-State-based data communities.
- Realizing an effective, coordinated national data ecosystem will require significant experimentation and innovation, and efforts to learn from them. The ability to learn lessons from successes and failures and to apply them in future interventions will determine the type of data communities and ecosystems, and the degree to which the continent harnesses the full transformative power of the data revolution.
- The importance of capacity needs assessments in terms of data requirements for the Sustainable Development Goals and Agenda 2063 and individual country readiness for different types of data and platforms.

References

Adieno, D. (2016). Citizen-generated data and the Sustainable Development Goals. Presentation at the Global Partnership for Sustainable Development Data Sustainable Senegal National Workshop on Data for Sustainable Development. Available from http://ipar.sn/IMG/pdf/civicus-datashift_-_citizen_generated_data_-_davis_adieno.pdf.

Alais, O. (2015). The impact of a move towards open data in West Africa. *Georgetown Journal of International Affairs*, 25 November 2015.

Blom, A., G. Lan and M. Adil (2016). *Sub-Saharan African Science, Technology, Engineering, and Mathematics Research: A Decade of Development*. Washington, D.C., World Bank.

Bonney, R. and others (2016). Can citizen science enhance public understanding of science? *Public Understanding of Science*, vol. 25, No. 1.

Boyera, S. and C. Iglesias 2014). Open data in developing countries: State of the art. Available from https://docs.google.com/document/d/1FMyfLu-jouL7j7Pw0kEwUn_B07aZ9IX3vlFGqPO0gX0/edit#!

Brynjolfsson, E., Hitt, L. M. and Kim, H. H. (2011). Strength in numbers: How does data-driven decision-making affect firm performance? Available from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1968725_code1376648.pdf?abstractid=1819486&mirid=1.

CEI (2016). DataShift in 2016. Available from http://cepei.org/wp-content/uploads/2016/11/DataShift_in_2016.pdf. Accessed on 23 February 2017.

Civicus (2017). DataShift: Building the capacity and confidence of civil society organisations to produce and use citizen-generated data. Available from http://civicus.org/thedatashift/. Accessed on 23 February 2017.

Communication Initiative Network (2016). Generating real-time citizen data to identify socioeconomic and political trends in East Africa. Blog. Available from http://www.comminit.com/global/content/generating-real-time-citizen-data-identify-socio-economic-political-trends-east-africa, Accessed on 6 December 2016.

Connelly, R. and others (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, vol. 59.

Davies, T., F. Perini and J.M. Alonso (2013). Researching the emerging impacts of open data: ODDC conceptual framework. ODDC Working papers No. 1. World Wide Web Foundation. Available from http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf.

de Montjoye, Y.-A. and others (2014). D4D-Senegal: The Second Mobile Phone Data for Development Challenge. Available at https://arxiv.org/pdf/1407.4885.pdf.

Einav, L. and J. Levin (2013). The data revolution and economic analysis. National Bureau of Economic Research. Working Paper No. 19035 Available from http://www.nber.org/papers/w19035.

Einav, L. and J. Levin (2014). Economics in the age of big data. *Science*, vol. 346, No. 6210.

EMC Corporation (2014). The digital universe of opportunities: Rich data and the increasing

value of the Internet of Things. Available from https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf.

Fisseha, E. (2012). The origin of national income accounting practices in Africa. *African Statistical Newsletter*, vol. 6, No. 1.

Fourie, J. (2015). The data revolution in African economic history. Available from http://www.wur.nl/upload_mm/b/1/5/a4bcd603-51c6-4e7b-ad38-8ea310c820c9_PAPER-The%20data%20revolution%20in%20African%20economic%20history%20-%20Johan%20Fourie.pdf.

Francescon, D. (2017). Research without borders: Sharing expertise in Africa. *Elsevier Connect,* 13 January 2017. Available from http://www.elsevierfoundation.org/research-without-borders-sharing-expertise-in-africa/. Accessed on 21 February 2017.

Frankema, E. and M. van Waijenburg (2013). Endogenous colonial institutions: Lessons from fiscal capacity-building in British and French Africa, 1880-1940. African Economic History Network, Working Paper Series, No. 11/2013. Available from https://www.aehnetwork.org/wp-content/uploads/2016/01/AEHN-WP-11.pdf.

Gervais, R. and R. Marcoux (1993). Saving Francophone Africa's statistical past. *History in Africa*, vol 20.

Goldman Sachs Asset Management (2015). The data revolution: From volume to value. Strategic Advisory Solutions, September 2015. Available from https://www.gsam.com/content/dam/gsam/pdfs/us/en/articles/2015/data-revolution-whitepaper.pdf?sa=n&rd=n.

Grace, K. and others (2012). Child malnutrition and climate in sub-Saharan Africa: An analysis of recent trends in Kenya. *Applied Geography*, vol. 35, Nos. 1-2.

Hilbert, M. (2016). Big data for development: A review of promises and challenges. *Development Policy Review*, vol. 34, No. 1.

Hlatshwayo, G. (2016). Swaziland data ecosystem mapping report. Background paper prepared for the Government of Swaziland and UNDP (unpublished).

IBM (2016). Outthink urban planning. Video clip. Available from https://www.ibm.com/events/ke/en/ted-outthink-urban.html.

Jerven, M. (2013). *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. Ithaca, Cornell University Press.

Jerven, M. (2014). Benefits and costs of the Data for Development targets for the post-2015 development agenda. Data for Development assessment paper. Copenhagen Consensus Center. Available from http://www.copenhagenconsensus.com/sites/default/files/data_assessment_-_jerven.pdf.

Jesuit Hakimaki Centre (2014). Open government data for effective public participation: Findings of a case study research investigating the Kenya's Open Data Initiative in urban slums and rural settlements. Available from http://oddc.opendataresearch.org/sites/default/files/publications/JHC%20Publication%20April%202014%20-%20ODDC%20research.pdf.

Kiregyera, Ben (2015). *The Emerging Data Revolution in Africa: Strengthening the Statistics, Policy and Decision-making Chain*. Stellenbosch, SUN Press.

Klosowski, T. (2016). Turn-by-turn navigation showdown: Google Maps vs. Waze. Available from http://lifehacker.com/turn-by-turn-navigation-showdown-google-maps-vs-waze-1761550298. Accessed on 13 November 2016.

Krätke, F. and B. Byiers (2014). The political economy of official statistics: Implications for the data revolution in sub-Saharan Africa. European Centre for Development Policy Management discussion paper 170. Available from http://ecdpm.org/publications/political-economy-official-statistics-implications-data-revolution-sub-saharan-africa/.

Kshetri, N. (2014). The emerging role of big data in key development issues: Opportunities, challenges, and concerns. *Big Data and Society*, vol. 1, No. 2.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META Group. File: 949. Available from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Lehohla, P. (2008). Statistical development in Africa in the context of the global statistical system. Paper prepared for the thirty-ninth session of the United Nations Statistical Commission. Available from https://unstats.un.org/unsd/statcom/doc08/BG-AfricaStatDev.pdf.

Lenard, T. and P. Rubin (2013). The big data revolution: Privacy considerations. Technology Policy Institute. Available from https://techpolicyinstitute.org/wp-content/uploads/2013/12/the-big-data-revolution-privac-2007594.pdf.

Lipton, M. (2013). Africa's national accounts mess. *Journal of Development Studies*, vol. 49, No. 12.

Long, J. and W. Brindley (2013). The role of big data and analytics in the developing world. Accenture. Available from https://www.accenture.com/us-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf.

Maaroof, A. (2015). Big data and the 2030 Agenda for Sustainable Development. Final draft report prepared for the United Nations Economic and Social Commission for Asia and the Pacific. Available from http://www.unescap.org/sites/default/files/Final%20Draft_%20stock-taking%20report_For%20Comment_301115.pdf.

Manyika, J. and others (2011). Big data: The next frontier for innovation, competition and productivity. McKinsey and Company. Available from www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx.

Manyika, J. and others (2013). Open data: Unlocking innovation and performance with liquid information. McKinsey and Company. Available from http://assets.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information.

Mdadila, K. (2016). Tanzania data ecosystem report. Background paper prepared for this report.

Miller-Rushing, A., R. Primack and R. Bonney (2012). The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, vol. 10, No. 6.

Mo Ibrahim Foundation (2016). Strength in numbers: Africa's data revolution. Available from http://s.mo.ibrahim.foundation/u/2016/05/16162558/Strength-in-Numbers.pdf.

Mungai, P. (2016). Kenya country data report. Background paper prepared for this report.

Nzayisenga, C. (2016). Rwanda national data ecosystem. Background paper prepared for this report.

Ochieng, C. (2014). Africa rising? Not without science, technology and innovation. *The African Technopolitan*, vol. 1 (September).

Ochieng, C. (2016). The poverty of development strategy in Africa. *The African Technopolitan*, vol. 5 (January).

Open Knowledge International (2016). Open data handbook. Available from http://opendatahandbook.org/. Accessed on 21 February 2017.

Organisation for Economic Co-operation and Development (2002). Measuring the non-observed economy: A handbook. Available from https://www.oecd.org/std/na/1963116.pdf.

Osamuyi, O. (2016). Google just turned on live traffic alerts for Maps users in Kenya, Nigeria And South Africa. Available from http://techcabal.com/2016/04/15/google-maps-live-traffic-alerts-nigeria-kenya-south-africa. Accessed on 24 February 2017.

Pricope, N. G. and others (2013). The climate-population nexus in the East African Horn: Emerging degradation trends in rangeland and pastoral livelihood zones. *Global Environmental Change*, vol. 23, No. 6.

Rakotomanana, E. (2016). Madagascar country report. Background paper prepared for this report.

Royal Statistical Society (2014). The Data Manifesto. Available from http://www.rss.org.uk/Images/PDF/influencing-change/rss-data-manifesto-2014.pdf.

Selassie, S. (2016). Ethiopia country data report. Background paper prepared for this report.

Theobald, E. and others (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, vol. 181.

Third International Open Data Conference (2015). Enabling the data revolution: An international open data roadmap – Conference report. Available from http://1a9vrva76sx19qtvg1ddvt6f.wpengine.netdna-cdn.com/wp-content/uploads/2015/11/opendatacon-report-en-web.pdf.

Uganda. Ministry of Finance, Planning and Economic Development (2012). Poverty Status Report: Poverty Reduction and the National Development Process. Available from http://www.trademarksa.org/sites/default/files/publications/2012-05%20Poverty%20Status%20Report_%20Poverty%20Reduction%20and%20the%20National%20Development%20Process%20-%20Uganda.pdf.

United Nations (2014). A world that counts: Mobilising the data revolution for sustainable development. Report of the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. Available from http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

United Nations, Economic Commission for Africa (2015). Africa Data Consensus. Available from http://www.uneca.org/sites/default/files/PageAttachments/final_adc_-_english.pdf.

United Nations, Statistics Division (2003). Handbook of Statistical Organization. Third edition. Available from https://unstats.un.org/unsd/dnss/hb/default.aspx.

United Nations, Statistics Division (2013). Fundamental Principles of Official Statistics. Available from https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.

United Nations Educational, Scientific and Cultural Organization (2016). UNESCO Science Report: Towards 2030. Second revised edition. Paris, UNESCO Publishing.

United Nations Global Pulse (2012). Big data for development: Challenges & opportunities. Available from http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf.

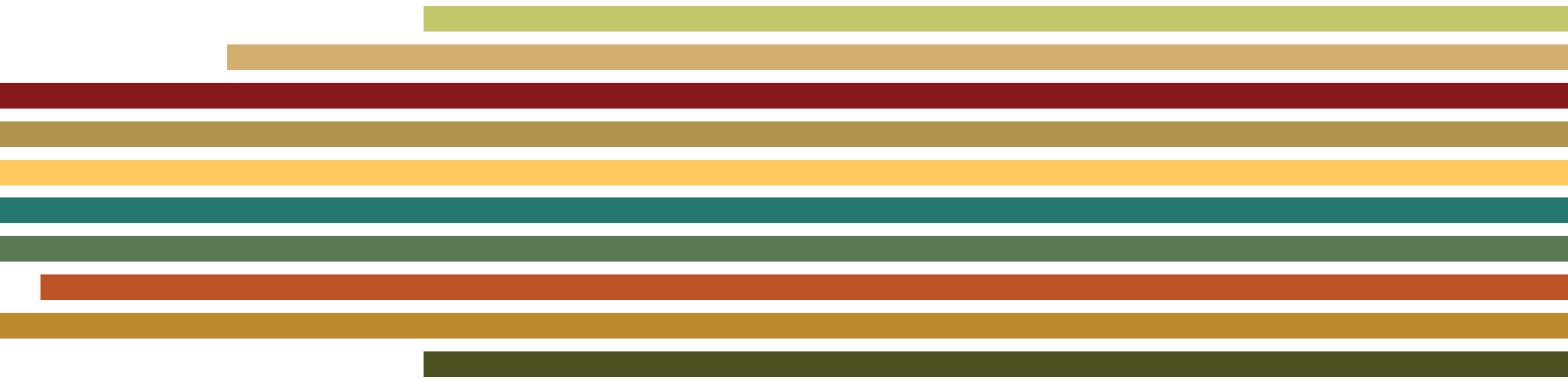
United Republic of Tanzania, National Bureau of Statistics (2010). Dissemination and pricing policy. Available from http://www.nbs.go.tz/nbstz/documents/Dissemination_and_Pricing_policy220kb.pdf.

United Republic of Tanzania, Ofisi ya Rais (2015). Speech by His Excellency Dr. Jakaya Mrisho Kikwete, the President of the United Republic of Tanzania during opening of Africa Open Data Conference held at Julius Nyerere International Convention Centre, Dar es Salaam 4th September, 2015. Available from http://www.ikulu.go.tz/index.php/media/speech/1809. Accessed 17 February 2017.

van Belle, J-P. (2016). South Africa country data report. Background paper prepared for this report.

World Bank (2014). Open data for economic growth. Available from https://openknowledge.worldbank.org/handle/10986/19997.

World Bank (2016). Open data and sustainable development. Policy note ICT01. Available from http://pubdocs.worldbank.org/en/741081441230716917/Open-Data-for-Sustainable-development-PN-FINAL-ONLINE-September1.pdf.



Printed in Addis Ababa, Ethiopia by the ECA Printing and Publishing Unit.
ISO 14001:2004 certified. Printed on chlorine free paper.