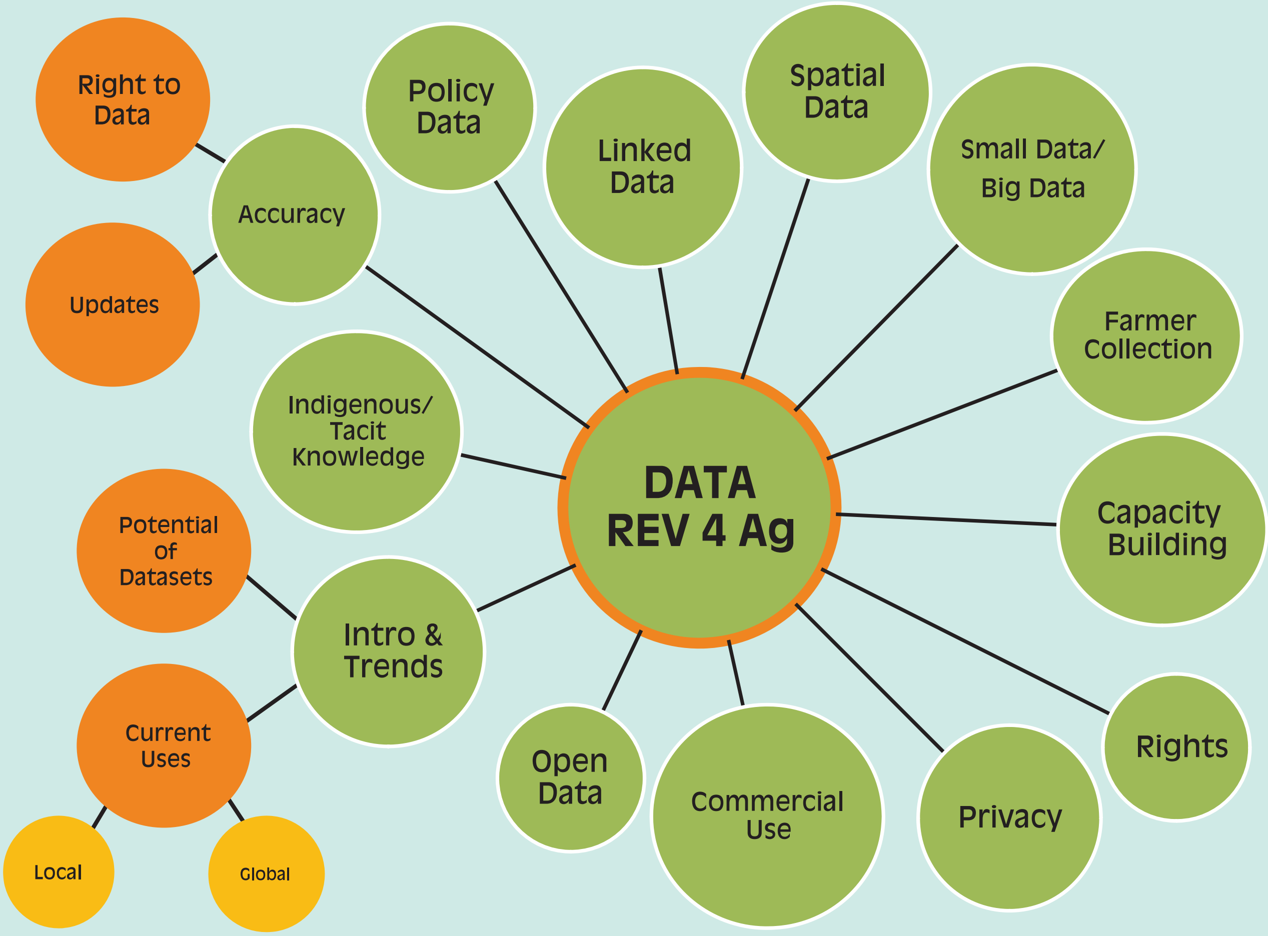
CTA Discussion Paper

Data Revolution for Agriculture



CTA
GODAN
Global Open Data for Agriculture and Nutrition
PAFO
Panafrican farmers' organization
Organisation panafricaine des agriculteurs
المنظمة الإفريقية للفلاحين

ABOUT CTA

The Technical Centre for Agricultural and Rural Cooperation (CTA) is a joint international institution of the African, Caribbean and Pacific (ACP) Group of States and the European Union (EU). Its mission is to advance food and nutritional security, increase prosperity and encourage sound natural resource management in ACP countries. It provides access to information and knowledge, facilitates policy dialogue and strengthens the capacity of agricultural and rural development institutions and communities.

CTA operates under the framework of the Cotonou Agreement and is funded by the EU.
For more information on CTA, visit www.cta.int

ACKNOWLEDGEMENTS

Contributors: Chris Addison (CTA), Isolina Boto (CTA), Ana Brandesescu (GODAN), Hugo Besemer (Wageningen University), Jerven Morten, Ben Schaap (GODAN), Isaura Lopes Ramos (CTA)

Editorial project management: Isaura Lopes Ramos
Editing: Susanna Cartmell-Thorp

Table of contents

GODAN

André Laperrière, Executive Director, GODAN Secretariat

In *Data Revolution for Agriculture*, CTA makes a powerful case for the empowerment of small-scale farmers and emerging economies through information.

With the emergence of low-cost, readily available open data, knowledge is no more the privilege for a few, but a right for everyone.

McKinsey recently referred to the data revolution in this context, crafting the concept of 'globalisation for the little guy', referring to data being no more the exclusive property of large corporations but rather, a tool that can enable everyone to tap into global intellectual capital.

Yet, with the emergence of low-cost technologies the world is becoming flooded – some would say cluttered – with information whose degree of relevance is often questioned. The true data revolution does not simply lie in floodgates of data suddenly opening up: it lies in concrete applications that make sense and use of it. This document is a reminder that empowerment and growth can only be achieved putting the beneficiaries in the driving seat, becoming masters of their decisions through enlightened and informed decisions. Based on its extensive hands-on experience working with small and medium-scale farmers in the most remote corners of the world, CTA describes with passion how information can transform the lives of rural populations, stimulate economic growth and in turn, lead the world ahead of the food security challenges ahead of it. The quest for open data described in these chapters demonstrate not simply how the world will better feed itself, but how everyone's quality of life will be improved in the process.

CTA's approach to open data is one of an advocate, but a critical one, emphasising the need for wisdom and discernment. For a sustainable and most beneficial development of our collective agricultural future, we must together strive to make sound and relevant data available, understood and put to use. That is the challenge proposed in the *Data Revolution for Agriculture*.

Pan African Farmers' Organization

Theo de Jager, President, Southern African Confederation of Agricultural Unions and Pan African Farmers' Organization

In the very same way as mobile communication technology has changed the life of a farmer, making it so much easier to run his/her business, big data is bound to irreversibly change the way we produce food and fibre in the next few years. More than a handy management tool, a source of information or a support mechanism for decision-making, big data will in months and years to come, become the air we breathe in our farming enterprises. We will, in a decade from now, wonder how we ever got along without real-time processed data at our fingertips, very much as we now wonder how we ever got along without mobile phones.

But for farmers and their organisations, there is one stark difference in the way mobile communication technology and big data technology is changing our enterprises, and that has less to do with the technology and more with ourselves. Somehow mobile technology caught us by surprise. We did not see it coming. What a pleasant surprise it was when we discovered the accessibility and applicability of these first brick size devices which could, in real time, connect us with anyone at any place in the world, any time! Soon data, cameras, global positioning systems and internet connectivity followed, and a farmer's life would never be the same again! Before we knew, it engulfed us, and we had a hard time keeping up with all the applications!

With big data, that is different. Thanks to the interventions of partners and sympathisers such

as CTA, GODAN and Bill and Melinda Gates Foundation, both multinational and national farmers' organisations know that the technological tsunami of big data is on its way, and we can prepare. Moreover, we can actually position ourselves and our structures to benefit to the maximum. We can ensure that we retain ownership of our data from the agricultural sector, and share in the business opportunities associated to the big data markets. In and through our farmers' organisations we can aggregate the sources of data like no one else can, and use that capacity to financially strengthen our structures to the point where it can become independent.

That's why in PAFO, SACAU, EAFF and the other regional farmers' organisations in Africa, we don't wait for the technology to reach us. We go out to seek and capture it, to embrace it and own it. Recognising the technological leapfrogging it will bring about, we as farmers know we have but this one chance to saddle it up and to ensure that we will keep on its back!

CTA

Michael Hailu, Director, CTA

We are living in the middle of an information and communications revolution. More than 40% of the worlds' population has access to the internet and nearly 7 out of 10 have a mobile phone. However, this means that 4 billion people still do not have internet access and nearly 2 billion do not use mobile phones.

If we are to take advantage of this digital revolution to work towards a more sustainable and inclusive world, developing countries must invest more in digital infrastructure and promote policies directed towards universal access that can benefit rural communities.

Access to timely and relevant data and information can play a key role in empowering rural communities, and opening access to data is key to this. Open access to data can accelerate food production and nutrition through supporting innovation in agriculture and value-chain development. Current estimates of the economic potential of open data range from hundreds of billions to trillions of dollars per year. At the same time, opening access to big data – call logs, mobile-banking transactions, online user-generated content such as blog posts and tweets, online searches, satellite images, etc. – and turning it into actionable knowledge can generate tremendous benefits for value-chain actors, in particular smallholder farmers.

As a knowledge broker with experience across African, Caribbean and Pacific countries, CTA is uniquely positioned to promote awareness of the benefits of open data based on its knowledge of tangible experiences in various contexts.

While others have focused on research information, CTA tries to look at a broad range of data that can benefit smallholder farmers and agribusinesses. We focus on two ways that these data can be leveraged to transform agriculture:

1. At the farmer and farmer-organisation level, to provide more information to support decision-making, including choice of crops, land management and development of more efficient value chains with better access to markets and greater availability of inputs
2. At the policy level, to support better decision-making and monitoring of progress.

But improving the quantity and quality of open data is not the only necessity; we must also make it more readily accessible and useful. To achieve this, we need to encourage better cataloguing and promotion of data resources by data owners and publishers. Adoption of clear open data policies, standards and platforms is key in promoting data sharing and utilisation across different domains.

We urge readers to support this data revolution by opening access to their data and actively seeking ways in which they can build innovations on this foundation to the benefit of people throughout the world.

"… the real challenge lies in ensuring that access to quality data is widely available and linked to local solutions for improving food security and nutrition."

Executive summary

New technologies are leading to an exponential increase in the volume and types of available data, offering possibilities and innovative solutions for all global societies. This represents an opportunity – for instance to target tailored interventions to a specific group – as well a challenge, particularly in terms of privacy rights and potential abuse. But it is not just a matter of creating more data; it is about making the best use of up-to-date and readily accessible data.

Climate change and population growth are having major impacts on agriculture and food security. Rural communities who are heavily reliant on agriculture as a critical source of livelihoods and employment are expected to suffer the most from these stress factors. The use of reliable, timely and accessible agricultural and nutritionally relevant data can help in coping with these challenges.

In terms of agriculture and open data, information is a key element for achieving food security and improved nutrition.

Although in recent years, open data[1] has been steadily climbing up the global agenda and the amount of data available is constantly increasing (for example, through the use of satellite imagery and mobile technology), there are still challenges constraining open agricultural data in African, Caribbean and Pacific (ACP) countries. This often results in inaccurate, outdated, incomplete or inaccessible data due, for instance, to costs related to statistical surveys, weak capacities, methodologies and tools and insufficient involvement of some key agriculture and rural development actors.

Providing smallholder farmers with accessible and accurate data can translate into an increase in agricultural productivity and improved nutrition. However, the real challenge lies in ensuring that access to quality data is widely available and linked to local solutions for improving food security and nutrition.

The main objective of this publication is to demonstrate the impact and the use of open data for agriculture by showcasing applications, specific trends, approaches, case studies and highlighting the main challenges in this area. When promoting the use of open data in agriculture and rural development, it is important to adapt a driven, inclusive process allowing the full engagement of local communities and the creation of new businesses for a broad range of stakeholders.

This publication has been produced jointly by the Technical Centre for Agricultural and Rural Cooperation (CTA) and the Global Open Data for Agriculture and Nutrition (GODAN) with contributions from the Pan African Farmers' Organization. It encompasses a series of studies, reports, and background documents previously produced by CTA on the potential impact and use of agricultural open-data policies and practices across the ACP region.

This paper highlights two key areas:

- Global initiatives and partnerships to improve data collection and use
- Use of open, big data in agriculture as a complementary for poverty reduction.

Some key concepts

Open data. Various definitions for open data exist. According to Opendefinition.org, "Open data and content can be freely used, modified and shared by anyone for any purpose". It means that data are:

- Available and accessible: the data must be available as a whole, and at no more than a reasonable reproduction cost. The data must also be available in a convenient and modifiable form.
- Reused and redistributed: the data must be provided under terms that permit reuse and redistribution, including the intermixing with other datasets.
- Universal participation: everyone must be able to use, reuse and redistribute. There should be no discrimination against fields of endeavour or against persons or groups.
- Reliable: meaning that the data has guaranteed availability and consistency over time, so that others can rely on it.

1 Open data is defined as data that anyone can access. More at http://opendatahandbook.org/guide/en/what-is-open-data/

Closed data are data that should not be shared. Private data about individuals falls into this category. Many valuable datasets are derived from personal data. While it may be possible for aggregated and derived data to be shared, care must be taken to ensure that individual rights to privacy and control over their own data are not undermined, and that personal data are managed according to international best practices.

Shared data are data that are made available to specific users, or for particular kinds of reuse. For example, a telecom firm may give government or researchers access to anonymised phone records to allow them to understand social trends, or government may share aggregated health records with pharmaceutical firms to support medical research.

Data revolution. The 'data revolution' was first coined in May 2013 in the report of the High-Level Panel on the Post-2015 Development Agenda. It defines the explosion in data volume, the speed with which data are produced, the number of data producers, data dissemination, and the range of data coming from: new technologies such as mobile phones; the 'internet of things' (connecting devices over the internet); and other sources, such as qualitative data, citizen-generated data and perceptions data.

Real-time digital data. The growing role of 'crowdsourcing' and other 'participatory sensing' efforts bring together communities of practice and like-minded individuals through the use of mobile phones and other platforms including the internet, handheld radio and geospatial technologies etc.

Big data. Advances in computing and data science now make it possible to process and analyse 'big data' in real time. However, due to its size and often complex and unstructured nature, big data presents several analytical challenges that demand continually updated tools and expertise.

Big data is a popular phrase used to describe a massive volume of both structured and unstructured data that is so large that it is difficult to process with traditional database and software techniques (Addison *et al.* 2015). The characteristics which broadly distinguish big data are also called the "**Five Vs": volume, velocity, variety, veracity and value**.

Volume refers to the vast amounts of data generated every second; just think of all the emails, Twitter messages, photos, video clips, sensor data etc. produced and shared every second. On Facebook alone, 10 billion messages are sent each day, the 'like' button is clicked 4.5 billion times and 350 million new pictures uploaded every single day. Before long, the same amount of data will be generated every minute as all the data generated in the world between the beginning of time and 2008! This means that data sets too are increasingly becoming too large to store and analyse using traditional database technology. With big data technology, data sets are stored with the help of distributed systems, where parts of the data are stored in different locations and brought together for use by software.

Velocity refers to the speed at which new data are generated and moves around. Just think of social media messages going viral in seconds, the speed at which credit card transactions are checked for fraudulent activities, or the milliseconds it takes trading systems to analyse social media networks to pick up signals that trigger decisions to buy or sell shares. Big data technology also allows data to be analysed while it is being generated, without ever putting it into databases.

Variety refers to the different types of data that can now be used. In the past, the focus was on structured data that neatly fits into tables or relational databases, such as financial data (e.g. sales by product or region). In fact, 80% of the world's data are now unstructured, and therefore cannot easily be put into tables (e.g. photos, video sequences or social media updates). With big data technology, different data types (structured and unstructured) can now be harnessed including messages, social media conversations, photos, sensor data, video or voice recordings and brought together with more traditional, structured data.

Veracity refers to the messiness or trustworthiness of data. With many forms of big data, quality and accuracy are less controllable (think of Twitter posts with hash tags, abbreviations, typos and colloquial speech, as well as the reliability and accuracy of content) but big data and analytics technology now allow us to work with these data types.

Value. Having access to big data is all well and good but unless it can be turned into value, it is useless. So it can safely be argued that 'value' is the most important 'V' of big data. It is critical that businesses make a business case for any

attempt to collect and leverage big data; it is so easy to fall into the buzz trap and embark on big data initiatives without a clear understanding of costs and benefits.

Some people refer to 'seven Vs' and add **visualisation** (presenting data in a manner that is readable and accessible) and **variability** (data whose meaning is constantly changing).

Big data for development

Sources of **big data for development** generally share some or all of these features (United Nations Global Pulse, 2014a):

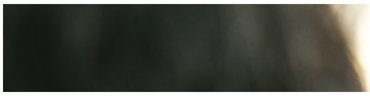
1. Digitally generated – i.e., data are created digitally (as opposed to being digitised manually), and can be stored using a series of ones and zeros, and thus can be manipulated by computers.
2. Passively produced – i.e., a product of our daily lives or interaction with digital services.
3. Automatically collected – i.e., there is a system in place that extracts and stores the relevant data as it is generated.
4. Geographically or temporally (time-related) trackable – i.e., mobile phone location data or call duration time.
5. Continuously analysed – i.e. information is relevant to human well-being and development and can be analysed in real time.

In some cases, big data is defined by the capacity to analyse a variety of mostly unstructured data sets from sources as diverse as weblogs, social media, mobile communications, sensors and financial transactions. This requires the capability to link data sets, which can be essential as information is highly context-dependent and, in the wrong context, may not be of value. It also requires the capability to extract information from unstructured data, i.e., data that lack a predefined (explicit or implicit) model. Estimates suggest that the proportion of unstructured data in businesses could be as high as 80% to 85% and largely unexploited or underexploited. In the past, extracting value from unstructured data was labour-intensive. However, with big data analytics, unexploited data silos can be linked and analysed to extract potentially valuable information in an automated, cost-effective way.

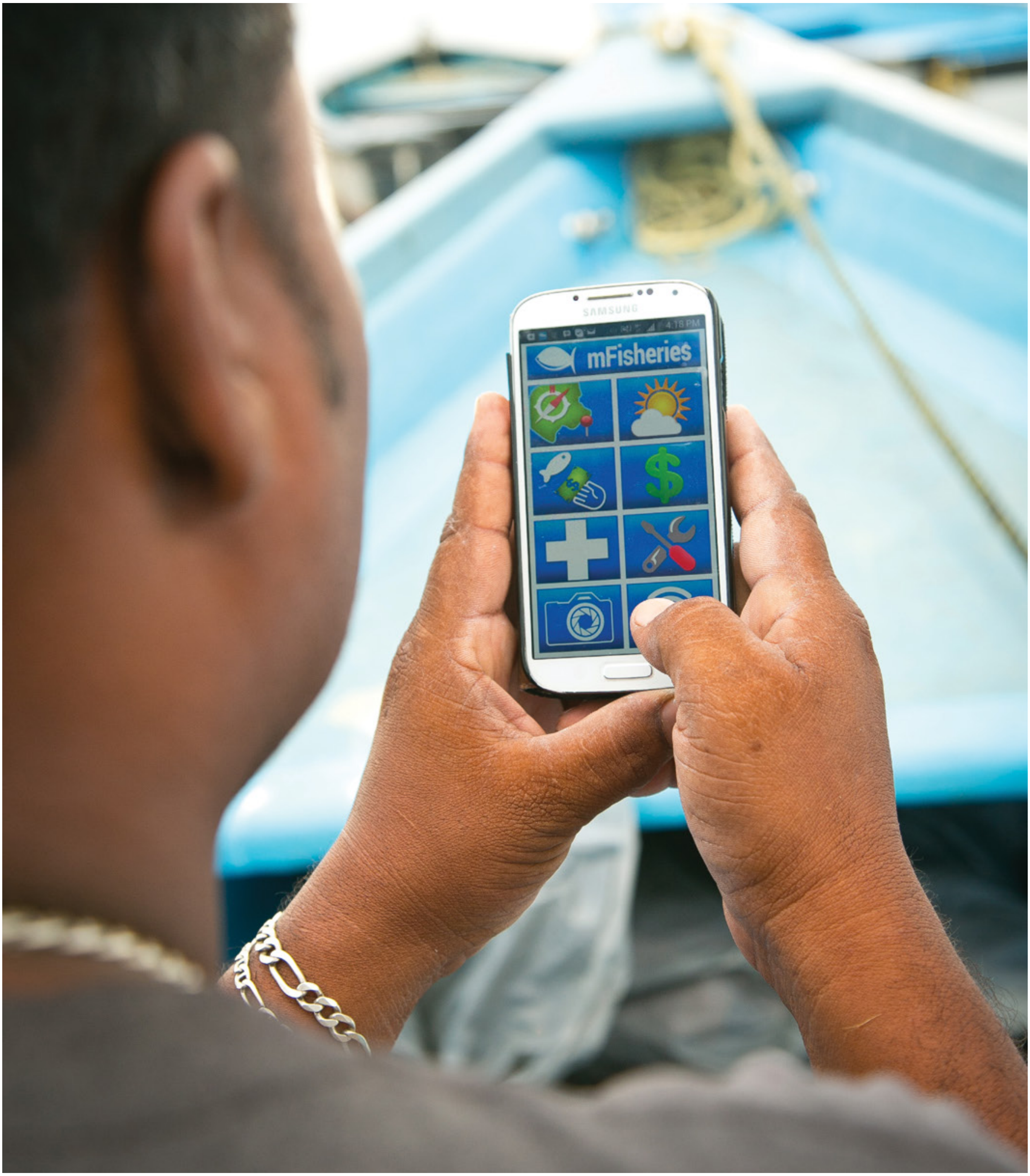
Delivering value from big data is a challenge. Overcoming this requires building capacity in three distinct areas, namely: i) scalable data management (processing, storage, resource management); (ii) data analysis (harnessing statistics generated); and (iii) expertise in the field, i.e. sustainable agricultural development.

Data blending is a quick and straightforward method used to extract value from multiple data sources. This process can also help to discover correlations between different data sets without the time and expense of traditional data warehouse processes. For instance at CTA, the documents' collection is linked to the Food and Agriculture Organization (FAO) Agrovoc system[2] for keywords and geocodes for places to help people navigate through CTA's collection of nearly 50,000 documents online.

2 http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus



"The rise in prominence of open data has coincided with a renewed emphasis on the role of data and information in international development ... and advances in ICTs"



CHAPTER 1

Introduction

The open data concept

Open data rose to global prominence in 2009 when the then newly appointed President of the United States, Barack Obama, signed the US Open Government Directive on his first day in office. Since then, 'open government' has become one of the most significant public policy and technological trends around the world, culminating in September 2011 with the launch of the Open Government Partnership (OGP). The Open Government Partnership[3] is a multilateral initiative that aims to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption and harness new technologies to strengthen governance. To date, 67 governments joined the Partnership, and over 1,000 commitments were made for governments be more open and accountable.

Shortly after, the European Commission (EC) presented a comprehensive open data package looking at the potential for further opening up data, in particular, data produced by the public sector. One of the key elements of the package is to reinforce the main legal instrument applicable across the 28 EU Member States, the Directive 2003/98/EC on reuse of public-sector information (PSI Directive). As a rule, the EC has allowed reuse of its documents for commercial and non-commercial purposes at no charge since 2006. In 2011, the EC also committed to work towards providing documents in machine-readable format, where possible and appropriate, and to set up an open data portal to promote the accessibility and reuse of this information. This was followed, in December 2012, with the launch of the EU Open Data Portal.[4]

At the G8 Conference in London in October 2013, the Global Open Data in Agriculture and Nutrition initiative (GODAN) was launched (van Vark, 2013). GODAN's objective is to build high-level policy and institutional support for open data relevant to agriculture and nutrition across the public and private sector to enforce global food and nutrition security.[5]

To date, more than 250 governments at national, sub-national and city levels across almost 50 developed and developing countries, and entities such as the World Bank and United Nations have launched open data initiatives (World Bank, 2014).

The rise in prominence of open data has coincided with a renewed emphasis on the role of data and information in international development, which is largely a result of advances in information and communication technologies (ICTs), particularly the mobile phone. With more than 3.5 billion unique mobile subscribers worldwide, the mobile phone has not only become one of the most rapidly adopted technologies in history, but has also proved to be a significant tool for improving data collection.

The recognition of this value has been reflected in the High-Level Panel on the Post-2015 Development Agenda's call for a 'data revolution' for sustainable development, with greater focus on improving statistics and information capabilities to track progress, increase data-driven decision-making and strengthen accountability.

Why data matter

Good, available, reliable and well-used data can make a difference to people's lives. Data allow the most vulnerable groups to be reached, their specific needs to be tackled and progress tracked. However, data are also important to businesses and individuals who use information to make decisions daily that affect their well-being (SDSN, 2015).

Open access to reliable information allows citizens to acquire information and track different service provision that impact on their lives (e.g. information on education, healthcare, etc.), to hold governments to account and to better

3 http://www.opengovpartnership.org/
4 https://open-data.europa.eu/en/data/
5 http://www.godan.info/

Example data types

Census: Systematic recording of information from all members of a given population.

Household survey: National sample of randomly selected households that provides data on demographic and socioeconomic characteristics.

Agricultural survey: Surveys of farms and people who operate related enterprises, including data on crop yields, economic variables and environmental data.

Geospatial data: Data information about a physical object that can be represented by numerical values in a geographic coordinate system.

Administrative data: Information collected primarily for administrative or management purposes, including welfare, taxes, educational records, etc.

Economic statistics: Financial and economic performance measurements, including labour force and surveys, economic performance, employment, taxation, imports and exports, etc.

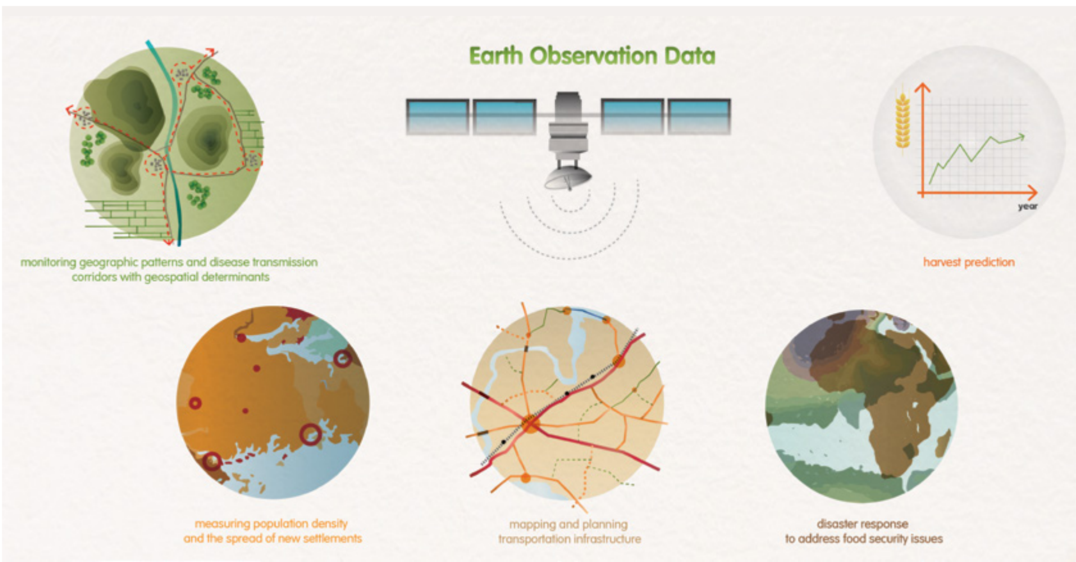
Environmental data: Real-time monitoring, ground stations and satellite imagery for a range of environmental variables including biodiversity, water resources quality, water resources and forest and land use.

Source: Adapted from SDSN (2015).

participate in democratic processes in order to ensure their needs and concerns are taken into account. Data also inform government policies by providing inputs for accurate planning.

Mobile devices, biometric data, and crowdsourced citizen reporting have already changed official data collection processes and the design of the programmes they monitor. The same can be applied for satellite imagery. The cost of high-resolution image acquisition is falling while image availability and capacity for automated processing are increasing. There are many applications for earth observation data: harvest prediction, disaster response to address food security issues; monitoring geographic patterns and disease transmission corridors with geospatial determinants; measuring population density and the spread of new settlements; and mapping and planning transportation infrastructure. FAO and Google are working together to make high-resolution satellite data an everyday tool in managing the world's natural resources (FAO, 2016).

Many surveys are now being conducted on digital mobile platforms (SDSN, 2015), which reduces the time and cost of data collection. It also improves accuracy, simplifies Geographic Information Systems (GIS) and image data collection, streamlines integration with other information streams, and opens up the possibility of incorporating micro-chip based sensors into survey processes. Innovation is not just about adopting new technologies; it is also about improving existing ones.



Source: CTA (2016)

Figure 1. The myriad of uses of Earth observation data.

From MDGs to SDGs: How the data revolution can ensure sustainable development

The Millennium Development Goals (MDGs) guided the global development agenda from 2000 to 2015. One of the key lessons from the MDGs is that a lack of reliable data can undermine governments' ability to set goals, optimise investment decisions, and measure progress (Ologunagba, 2015). The Sustainable Development Goals (SDGs), which are set to guide the next 15 years of global development efforts to 2030, need to be supported by informed and strong statistical systems that can measure and incentivise progress across all the goals.

According to the UN, in order to achieve the SDGs, integrated action on social, environmental and economic challenges is required with a focus on inclusive and participatory processes. In this regard, in October 2014, the Independent Expert Advisory Group on the Data Revolution for Sustainable Development (IEAG) highlighted the opportunities and challenges the world is facing to improve data for sustainable development.

In 2015, the Sustainable Development Solutions Network, together with Open Data Watch and other partners, carried out a cost assessment of the core statistical tools needed to measure sustainable development called *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development.* This study estimates that at least US$1 billion a year will be needed to maintain and upgrade the statistical systems of 77 of the world's poorest countries, as well as to ensure there is sufficient data collection to monitor the SDGs (SDSN, 2015). Half of this money will come from domestic resources, while the other half will need to be provided by international assistance.

The quantity and quality of data coming from national official sources has been steadily declining since the early 1980s, particularly in Africa (FAO, 2006). Factors contributing to this scenario include lack of financial resources and staff to collect data, inadequate technical tools, lack of institutional coordination, lack of integration of agricultural statistics into the national statistical system, etc. These challenges will require a comprehensive framework at national and international level in order to improve data collection and reporting.

The role of private, open and big data for agriculture

The role of 'big data' and 'open data' in the context of agricultural statistics in ACP countries is an emerging but growing one.

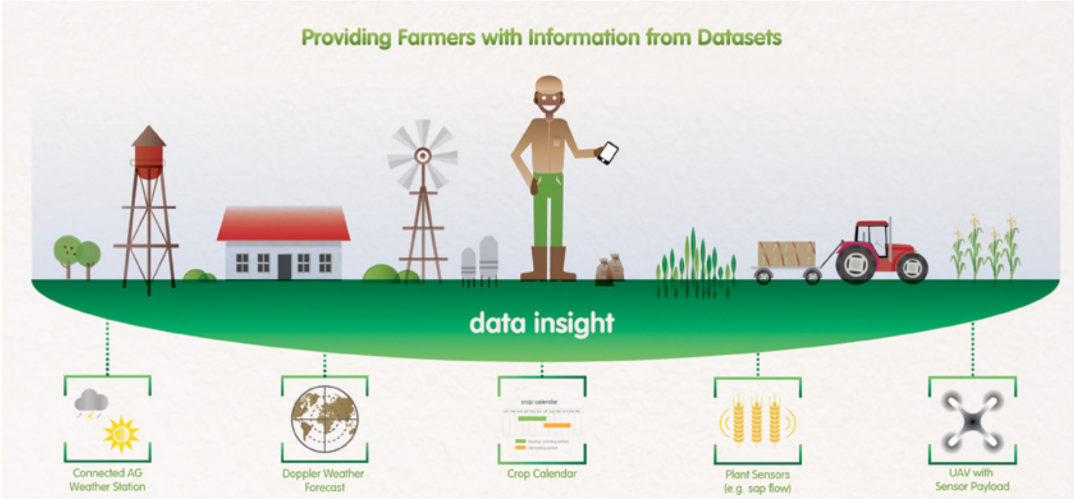
Big data sources provide huge data volumes which require storage and processing so using big data requires moving away from exclusive dependence on statistical methods that are unable to handle huge volumes of information. Instead, a more diverse set of tools should be adopted, which can be addressed through the use of data mining and machine learning algorithms with the required computational efficiency.

Furthermore, due to the enormous diversity of big data sources with regard to formats, content, storage, etc., methods to produce statistical information should be developed *ad hoc* for each case and the Generic Statistical Business Process Model might not be applicable. In the

The Namibia case: How the innovative use of data collection contributed to fight malaria

Namibia has made tremendous gains in malaria control and the epidemiological trend of the disease has changed significantly over recent years. In 2004, over 600,000 cases of malaria out of a population of two million people were recorded in Namibia. Seven years later, malaria cases had dropped to around 14,000, a reduction of 98%. Cell phone records were used to trace travel patterns, while satellite images helped researchers to draw detailed maps of vegetation, population density, rainfall and other drivers of mosquito and malaria parasite populations. Together, the two data types helped predict origin of infections and transmission from community to community. As a result, the Ministry of Health could target an estimated 80,000 people through the distribution of bed nets.

Source: Adapted from Tatem et al. (2014).



Source: CTA (2016)

Figure 2. Access to reliable data can translate for smallholder farmers into higher productivity, greater access to markets and better nutrition.

beginning, the big data source should be explored using data mining procedures to learn about the unknown data structure and decide on possible outcomes, and how to combine these data with traditional statistical production procedures (Addison *et al.*, 2015).

Generally, the majority of current programmes concerning the improvement of national statistic systems in developing countries, particularly in Africa, are focused on more traditional, orthodox data collection techniques. However new technology has made inroads in other areas of activity such that the information produced by technology in these areas has become increasingly important as a source of statistical data.

An example of this technology would be use of mobile telephony banking services, such as the mobile money transfer system M-PESA, which is widely used across a number of African countries. Operated by Safaricom in Kenya, M-PESA allows users to transfer money, pay bills and even withdraw cash from an automated teller machine (ATM), through their mobile phones, without having to register or qualify for a bank account (World Bank, 2011a).

In health, education, environment and even in security and disaster prevention sectors, other digital technologies are applied which contribute to the growth of information that could potentially be used by the national statistics office and other bodies responsible for agricultural statistics. Big data also raises important questions about the potential role of the private sector in national statistic systems, as the private sector continues to be a source of as well as a generator of big data.

As the global debate continues on protecting rights, particularly in terms of privacy, in the access to and application of big data in developing countries with large rural populations and capacity constraints, governments and civil society have not addressed the issue of public policy or regulation of big data with the same energy as in developed countries. Filling this governance vacuum over big data use will certainly become an issue that more stakeholders in ACP countries will have to pay attention to, particularly with a view to ensuring that the use of big data complements the overall SDGs and reduction of poverty which remain a priority in most ACP states.

CHAPTER 2

The potential impact of open data on agriculture

A brief overview

The current increased volume, velocity and variety of data used across the economy, and more importantly their greater social and economic value, represents a shift towards a data-driven socioeconomic model. Data can create a significant competitive advantage and drive innovation, sustainable growth and development.

Despite the key role of agriculture as an economic sector, serious weaknesses still persist in the measurement of agricultural outcomes. While governments and donors target agriculture for large-scale investments with ambitious goals of raising agricultural productivity, the monitoring of agricultural development is currently underpowered in order to produce good statistics.

Open data hold the potential – as yet largely untapped – to allow decision-makers to track development progress, improve social protection and understand where existing policies and programmes require adjustment. Whereas in previous generations, a relatively small volume of non-digital data was produced and made available through a limited number of channels, a massive amount of data is now regularly being generated and flowing from various sources, through different channels, every minute in today's digital age (Addison *et al.*, 2015). The challenge is to work jointly on a data agenda that can reap the benefits of the abundance of information.

Data for policy

Agricultural statistics continue to suffer from poor quality, lack of relevance, insufficient funding and are little used in national policy dialogues. The poorest countries – for which agriculture is a critical source of livelihoods – often have the poorest data in quality and scope, as they are least able to direct limited resources into improving statistical quality for informing policies (AfDB, 2013a). For instance, of the 44 countries in sub-Saharan Africa rated by the FAO, only two are considered to have high standards in data collection, while standards in 21 countries remain low (Carletto *et al.,* 2010).

Furthermore, serious concerns persist over growing populations linked to a rising global demand for food, fast-paced urbanisation, climate change, malnutrition and volatile food prices. Future demand for agricultural produce will thus put growing pressure on already scarce agricultural resources. According to FAO estimates, to feed a world's population of 9 billion in 2050, food production will need to increase by 70% (FAO, 2009). In this scenario, accurate agricultural information is essential for achieving sustainable agriculture and food security.

Sound information about land use and agriculture production data enable policy-makers to target appropriate and effective interventions. At the same time, providing farmers with datasets on access to market information and credit, real-time pricing and weather data, can contribute to ensure the proper functioning of food commodity markets.

Only 2 out of 44 countries in sub-Saharan Africa are considered to have high standards in data collection.

Data for agricultural development: evidence and policy[6]

Low-income countries suffer from key shortcomings when measured against international standards for the compilation and dissemination of official statistics. In the past decade, many social statistics have improved. Most countries have now been able to conduct a population census, whereas in earlier decades coverage was less complete. Nevertheless, key

6 This paper was originally prepared by Morten Jerven within the context of the Brussels Briefing on Data: the next revolution for agriculture in ACP countries? http://brusselsbriefings.net/past-briefings/40-data-the-next-revolution-for-agriculture-in-acp-countries/

Nearly one-fifth of African countries have not taken an industry survey to calculate GDP since 2000.

deficiencies remain especially in the provision of economic, labour and – in particular – agricultural statistics.

Recent events in Africa have highlighted this issue. Economic statistics regularly inform and support big headlines, but statistics themselves seldom make the news. This was not so when the Financial Times called the rebasing of gross domestic product (GDP) in Nigeria one of the biggest news stories coming out of Africa in 2014. On one level, the news of the jump in total GDP from Nigeria is good news – of course, it is good that Nigeria is richer than thought and it is a sign that the biggest African economy is working to improve data needed for economic governance. However, it also points to big knowledge problems and data gaps particularly as there was a lot of writing on political motives and the veracity of the new numbers from Nigeria (Jerven, 2014). The new numbers are indeed preliminary estimates, and have not yet been validated with data on the demand side of the economy (the new numbers are derived from estimating total output) and scholars will continue to struggle to square higher income and higher growth with continued high levels of poverty. The startling fact about the rebasing is that, technically, there is nothing wrong with Nigeria's GDP doubling from one day to another. However, whether the new number is too high or too low can be debated without any prospect of final agreement; GDP is always an approximation and any GDP estimate depends on a range of assumptions and data points that can all be contested.

For instance, all bathroom scales tell you the wrong total weight, by a few grams or by several kilos. This does not matter if you are interested in whether you are gaining or losing. The problem appears when you compare your own weight to your neighbours' weight or if the scales changed overnight and now tells a different total weight. That is what happened in Nigeria; they changed the scales. The previous system was using weights from 1990, whereas the new system is based on 2010 data. The same happened in Ghana when they changed from 1993 to 2006 as the benchmark year (Jerven and Ebo Duncan, 2012).

Nigeria had used weights from 1990 to estimate economic gains until 2014. How could it take almost 25 years before the data for economic governance was updated? The answer is a matter of priority. GDP numbers are a product of the national statistical system and many national statistical offices have been neglected since the 1980s as there were more pressing concerns than investing in reliable and regular data on economic activities.

Data for economic governance has been neglected for too long, so when the information is updated, there are bound to be a few surprises. The credibility of official statistics will benefit from a transparent and open process as the supply of data from African economies on the rise is meeting the demand of an increasing number of observers that are suddenly paying attention.

The state of the knowledge problem in agricultural statistics

"A country's GDP estimates are only as good as the data on which they are based. Although industrial production is believed to be rising sharply in most countries, nearly one-fifth of the respondent countries had not conducted an industry survey since 2000. Even fewer countries conduct regular surveys or censuses of agriculture, despite its criticality to the food security situation in the continent". This was the summary of the knowledge problem in agricultural statistics and how it pertains to measurement of economic production in a report prepared by the African Development Bank (AfDB, 2013b). According to metadata listed on the FAO agricultural bulletin board on data dissemination and quality of statistics only two African countries are considered to have 'high standards' in data collection, while standards in 21 countries remain 'low'. The situation is arguably not only because of neglect in investment in agricultural statistics; there have been efforts but not all have been successful. "Past investments and technical assistance efforts in the area of agricultural statistics have failed to produce sustainable systems, while existing statistics continue to suffer from poor quality, lack of relevance and are little use in national policy dialogues. A recent survey of agricultural statistics in Africa concluded that, "The existing lack of consensus on protocols and standards has resulted in agricultural statistics that suffer from uncertain quality, poor comparability and low credibility" (Carletto *et al.,* 2015).

There are three main sources of data that feed into the agricultural statistical system:

- agricultural census
- agriculture surveys
- administrative data

The agricultural census is mainly a mapping and stock-taking exercise. It provides a useful benchmark for future surveys and administrative data, and its direct output is to build the capacity to provide better data in the future. Conducting a census is expensive, and if the sector is in rapid change, the data may quickly become outdated. FAO recommends one census every decade, but they are expensive and complex to organise. The number of countries in Africa taking the agricultural census has been small – 1 in 1930, 3 in 1950, 16 in 1960, 22 in 1970, 17 in 1980 and 14 in the 1990 round of the World Census of Agriculture Programme. Only Ethiopia had been able to do a census twice.

A cheaper alternative is to conduct sample surveys which yield production data. However, data are not available soon enough for policy-makers to make direct use of them. Furthermore, surveys rely on recall and studies using diaries to record production and consumption show that recall is weak. According to FAO and AfDB, over the past two decades about 50% of African countries have conducted either a survey or a census, and most of these countries have opted for doing a census instead of a survey (FAO and AfDB, 2012). In other words, most data received are from administrative sources – which are based on what officials working within the Ministry of Agriculture are reporting. Thus one way forward would be how to improve the reporting of administrative data.

Knowledge and governance problems

It is instructive to think about these as issues of knowledge and governance. Knowledge questions relate to scholarly pursuits of studying what works and what does not, and aims to unearth causal effects in one country, time or place that can be of relevance to further academic knowledge, or to further lessons that can be useful for policy-makers.

Governance questions are the ones that policy-makers want and need answers to and relate directly to current or future policy. We often think of a policy circle where data on a particular problem may inform or prompt a particular policy, and then data will again form an important role in providing feedback on the policy or a formal impact evaluation of the policy. Statistics matter because they help to identify needs, set goals and monitor progress.

A typical knowledge question could be whether fertilisers work. Scholars can provide an exact answer to this, and it does not require an agricultural census and a regular survey programme to be complemented to answer this question. Duflo *et al.* (2008) investigated whether fertilisers work for smallholder farmers. They conducted a randomised control trial with smallholders in Kenya where they gave some farmers no fertiliser, some too little, some just right and some too much. They could with remarkable accuracy conclude, “Fertilizer, when used in appropriate quantities, is highly profitable”. However, this is not the question policy-makers are facing. The political governance problem is whether a national fertiliser subsidy can be justified to donors and voters. One example is Malawi. Malawi had what was thought to be a successful fertiliser subsidy programme. It was celebrated by Jeffrey Sachs in New York Times: “Farm yields soared once nitrogen got back into the depleted soils” but, as documented (Jerven, 2014), while the production data for 2006/7 from the Ministry of Agriculture showed 3.4 million tonnes of maize, the Agricultural Census for 2006/7 census showed 2.1 million. The reason for the discrepancy was probably the vouchers for the fertiliser subsidies – the Ministry of Agriculture data showed about 1 million more farming households than the results from the census. This implies that either a large number of households registered themselves twice for the voucher programme or that the census reporting is flawed by under reporting of farmers and/or yield.

What about the data revolution in African agriculture?

There are certainly some quick wins for agricultural statistics derived from the use of technology. Most obviously, there are reduced costs and time savings in recording (global positioning system mapping) and reporting (connected handheld devices) and data processing. Land can now be surveyed without sticks and rope. However, the term ‘data revolution’ aims higher; it is not only about reduced costs and increased accuracy as expressed in a World Bank blog:

According to IBM, every day the world generates 2.5 quintillion bytes of electronic data.

> *Instead of writing large grants, spending days travelling to remote field sites, hiring and training enumerators, and dealing with inevitable survey hiccups, what if instead you could sit at home in your pyjamas and, with a few clicks of a mouse, download the data you needed to study the impacts of a particular program or intervention?*

The piece goes on to explain that satellite images can measure the reflectance from crop leaves that can be used to estimate yields from space. The caveats to these promising prospects are most telling. The technology is "most easily employable if the researcher already knows the plot boundaries and knows what crop is being grown".

Not to mention to apply the technology in cloud free areas to even be able to measure anything. Current technology will not able to capture very small effects and "satellite imagery was generally too coarse to resolve the very small plot sizes (e.g. less than half an acre) common in much of Africa." Finally, applying these approaches in low-income countries remains difficult "because of a lack of either (i) ground truth data to develop the satellite-based predictions, and/or (ii) a satisfactory mechanistic understanding in these environments" (Kondylis, 2015).

Moreover, even if yields and plot sizes were measurable from space, satellites will still not be able to solve major issues such as costs, income and prices. For some locations and activities, a bigger volume of data may be accessed through, for example, mobile phone data, but the samples remain unrepresentative since mobile phone penetration is still relatively poor among the rural poor.

Data for investment

One of the biggest impediments to investment is lack of information. With information, investor confidence grows. With confident investors, money flows and, sustainable development happens.

According to IBM, every day the world generates 2.5 quintillion bytes of electronic data (Wall, 2014). Big data is quickly being seen as the next big investment in agribusiness. For example, in November 2013, Monsanto Company completed its US$930 million cash purchase of the Climate Corp., a San Francisco-based tech company, which processes 52.43 million megabytes of weather information daily including eight years of soil, moisture and precipitation records for each of the 29 million farm fields in the U.S. Using this data, the company uses an algorithm that divides the country into approximately half a million plots and then generates 10,000 daily weather scenarios for each of them. This information is used to create custom operational advice to farmers and to create insurance policies for corn, soybean and wheat farmers. Then when data shows adverse weather conditions causing crop losses an insured farmer receives a cheque without the need for claims, forms or negotiating.

The potential impact of open data on the smallholder[7]

Smallholder farmers that trade their crop are one of the stakeholders in a larger ecosystem that consist of many value-chain actors including cooperatives, input providers, traders, processors, exporters and wholesalers, and global businesses (see Figure 3). Along the value chain there are different service providers: financial (credit, insurance), logistic (transport, storage, grading, certification) and extension (farm management advice, business advice) facilitating the value chain. There are also 'governance' actors such as the local, regional and national governments, donors, NGOs and researchers. Each of these actors is interested in information about production, functioning of the value chain and availability of services and governance. The more and better the information that is available about the smallholders' ecosystem and its functioning, the better the different actors can fulfil their role, ultimately strengthening smallholders' food and nutrition security. However, not all actors are interested in the same information or look at it in the same way. A smallholder wants to have the information needed for decision-making in the context of his or her farm: What crops should I grow? How do I grow these crops? Which inputs do I use and where can I get them? Where do I store my harvest? Where do I sell my crops

7 Jellema *et al.* (2015).

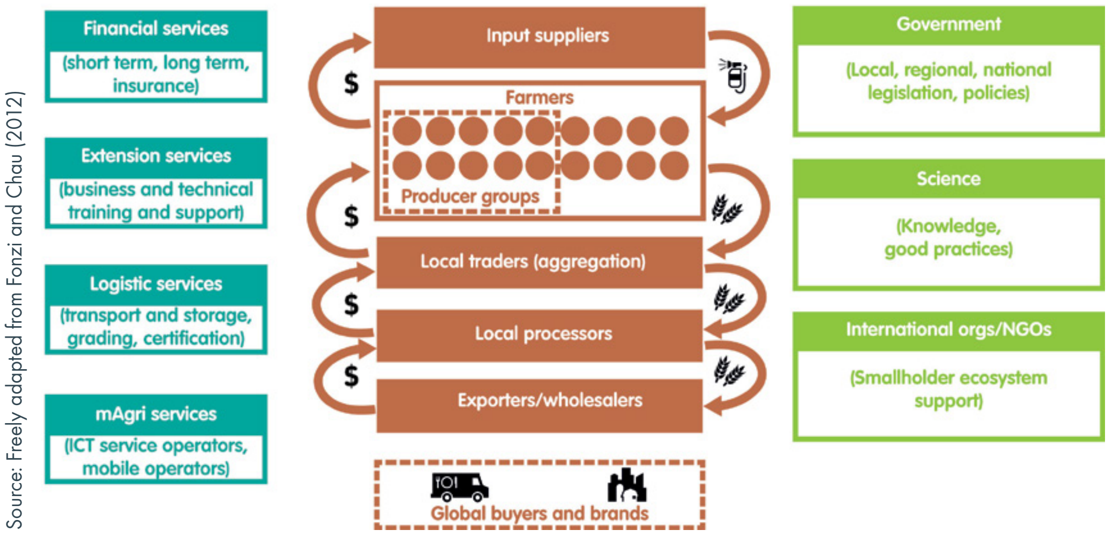


Figure 3. The smallholder ecosystem and its actors including the mAgri services.

and at which price? Whereas, a government may be more interested in the general picture of an area so it can adapt its policies accordingly. Ease of access to information in the ecosystem will ensure smooth functioning of the smallholder ecosystem. Sharing open data makes information more available in a transparent manner and can improve decision-making for multiple stakeholders.

Agricultural open data and ICT tools

Mobile operators and ICT service providers are the most recent emergent actors in the smallholder ecosystem. Sharing of data or information exchange between smallholders, or between smallholders and other actors, is traditionally very difficult in rural areas of developing countries because of large distances, bad roads and sparse, weak landlines for telecommunication. However, this picture has been changing in the last 10 years. ICT's affordability and universality has made them valid tools for data collection – resulting in more smallholder farmers connected to the outside world – and for obtaining data sets at a faster pace (World Bank, 2011a).

The mobile and ICT service operators are developing and hosting agricultural advice services on these mobile platforms, providing information as text messages, structured menus, voice messages etc. In some cases, these services are developed as part of the rural marketing strategy of mobile providers. The GSMA, an association of mobile operators, has currently identified 122 services deployed worldwide.[8]

Mobile agricultural advice services vary from: enhancing productivity on the farm; facilitating farmers' access to microcredit and insurance; helping cooperatives organise their stock and trade; allowing inclusion of farmers and cooperatives in agribusiness supply chains or providing access to the global market (World Bank, 2011a). In addition, farmers are also providing information about themselves and their environment through these services. Opening up these data streams will provide the opportunity to better understand smallholders and therefore the world at large, enabling better progress in terms of development and governance and achievement of the SDGs.

The following paragraphs describe three such tools.

iFormBuilder: An iOS mobile data collection platform that features an application that requires no paper or internet connection and is available worldwide. This application is being used for data collection in over 110 countries and it allows real-time data upload and offline data collection, while immediately sending any updates to a mobile workforce with server assignment. For example, Catholic Relief

8 http://www.gsma.com/mobilefordevelopment/tracker/

Services used iFormBuilder to register and distribute vouchers to beneficiaries during a seed fair in the Central African Republic, and they were able to save over a week of preliminary work and reduce staff by 50%.

Cropster: A Latin American initiative that seeks to support sustainable agriculture by empowering farmers with access to key information while ensuring data transparency. The application enables them to make informed decisions, and also supports people and communities at all supply chain levels. This app offers a monitoring and evaluation tool that facilitates data collection as well as the exchange of information within producer groups, NGOs and commercial partners. This tool provides decentralised monitoring, real-time information and the ability for users to customise data and verify input.

EpiSurveyor: An award winning mobile app that lets users create an account, design forms, download them to their mobile phones, collect data and send it to a server. According to a 2010 World Bank report nine data collectors used EpiSurveyor in 2010 to interview beneficiaries in 25 municipalities in a second survey (the first one, conducted in 2009, used paper and pen) in a World Bank Conditional Cash Transfer project in Guatemala. Digitisation cut the interview cost by 71%, increased sample size from 200 to 700 beneficiaries, and reduced individual interview time by 3.6%.

Implications for big data in agriculture: the leading role of the private sector and potential benefits for small-scale farmers

The relationship between the private sector and big data in agriculture, particularly in developing countries, is potentially the result of the private sector being the source of many key technological innovations, and subsequent diffusion of these innovations. Multinationals have played a strong role in the diffusion of new technologies which then generate big data, and the activities of enterprises – including exports, outsourcing, and licensing – have led to the rapid spread of new technologies in far flung countries, including those which may not necessarily be priority countries in the development agenda.

Furthermore, developing countries are increasingly developing technologies which are leading to the big data boom, often through collaborative relationships and in response to sectors that have been neglected during the technological boom.

An example is the AgriLife platform developed by Kenya-based IT company MobiPay and launched in late 2012.

However, a big challenge for agricultural big data, particularly in developing countries, is accessibility and, in some cases, affordability. Farmers, particularly smallholder farmers, are the poorest and most vulnerable stakeholders in the agricultural supply chain and, as such, are at a particular disadvantage in terms of being able to garner big data revolution benefits. Large growers can afford specialised machineries, which is not the case for small-scale farmers.

Big data is needed most by farmers who can least afford it. Major corporations are investing heavily in big data for agriculture, and start-ups are proliferating supported by the increasing availability of venture capital. But all this market-driven activity does little to help the poor in developing areas such as sub-Saharan Africa, where productivity is very low by European standards and where most of the world's population growth is predicted to take place in the coming decades. As big data increasingly plays a role in developed countries to increase productivity, the gap between developed and developing countries could become even greater and the perceived limited application of new technologies in farming in developing countries also decreases the attractiveness of the sector for investors and for the youth as a source of future employment.

Precision agriculture

In recent years, an increasing amount of geospatial data are being generated, from the use of maps via our mobile phones to global positioning system (GPS) driving directions leading to a fundamental shift in the way we are living and are informed.

A GIS captures, stores, analyses, manages, and presents geographic location data thereby allowing users to make decisions informed by highly accurate and detailed geographic information. Geospatial information applied

to agriculture plays a critical role in providing the right information to key decision-makers about the right practices to improve and to optimise food production. For instance, GIS can analyse soil data combined with historical farming practices to determine the best crops to plant, how they should be planted and how to maintain effective soil nutrition levels to achieve optimum productivity.

Precision agriculture (also known as precision farming, satellite farming or site-specific crop management) uses not only GIS and GPS but also remote sensing and variable-rate technologies to observe, measure and respond to spatial variations in crops, as well as monitor yields. The approach is rapidly changing the way farmers and agribusinesses are looking at crops and relate to the land. Precision data, together with computer-based decision support systems, help optimise production (yield), conserve resources (e.g. water and nutrients) and reduce costs (Venkatalakshmi and Devi, 2014). Examples of (satellite-based) crop monitoring services include Cropio, FarmSat, FieldLook and ClimatePro (Precision Agriculture, 2014). Data are also collected at all stages along the value chain, from farm to the consumer.

For many years, precision agriculture was considered irrelevant to small-scale farmers in developing countries because of the coarse resolution and high cost of the images. This has changed. There is now a growing body of research to support the idea that small-scale farmers can benefit from precision agriculture. GPS-equipped sensors on tractors, for example, enable farmers to measure and respond to soil variability across vast tracts of land, and dispense the right amounts of fertiliser and water exactly where it's needed.

Multilateral agencies, such as the World Bank and Asian Development Bank, have taken up a more focused approach towards use of spatial technologies and information for ensuring food security. The UN organisations are promoting Global Geospatial Information Management; the Famine Early Warning Systems Network, created in 1985 by the US Agency for International Development (USAID), is a leading provider of early warning and analysis on acute food insecurity; the Dutch government, through the Netherlands Space Office, has started a programme called Geospatial for Agriculture and Water. This initiative uses satellite data to improve food security and has projects in Bangladesh, Ethiopia, Indonesia, Kenya, Mali, as well as Uganda, where CTA, though the Market-led User-owned ICT4Ag Enabled Information Service (MUIIS) project, will use satellite generated data to improve production and marketing prospects for producers involved in maize, soya bean and sesame value chains.

The Copernicus Land Monitoring Service also makes use of satellite and *in situ* data to provide regular geospatial information on the state of global vegetation and water cycle for spatial planning, forest management, water management, agriculture and food security.

Land registration in Rwanda

10.3 million plots of all land in Rwanda were registered for the first time through the Land Tenure Regularisation Support Programme (LTRSP) – a two phase project, funded by UK Department for International Development (DFID), that involved a one-off, low-cost, community-based process of land-tenure regularisation.

LTRSP used the 'General Boundary Principle' to demarcate plots that uses natural landscape features to record approximate boundary positions – which can be revised if necessary – rather than formal surveying. Landowners were involved at each stage and were strongly encouraged to make themselves available for the demarcation process so they could guide para-surveyors and committees to the location and boundaries of their land, recording objections and disputes. Neighbours and communities were also involved when individuals' plots were being recorded to verify boundary accuracy.

LTRSP adopted a mix of commercial software and open source solutions. While the GIS unit used predominantly commercially licensed software, the Land Tenure Regularisation Support System and Land Administration Information System were developed solely using open source solutions. LTRSP was the first large-scale land registration programme to demonstrate the use of open source software for data processing, and is considered the first of its kind to use the technology systematically.

Participatory 3D modelling (P3DM) is a community-based tool – better defined as a 'process' – that integrates local spatial knowledge with data on land elevation and sea depth to produce a physical model.

Sample of sources of agricultural data[9]

Agricultural census enumeration areas. In many countries, cartographic materials and data from the population census are also used for the agricultural census. The sampling frame consists of enumeration areas and aggregated data from the census data collection. As in the population census, random samples of enumeration areas are selected and screened for farms or agricultural holdings for agricultural production surveys.

Farm registers from the agricultural census. As in the household registers, countries with adequate capacity can use the agricultural census to develop farm registers, which provides a powerful sampling tool because it allows a choice of many alternative sampling designs. A major weakness is that the registers rapidly become out-of-date; out-of-date population and farm registers erode all of the data quality dimensions because the completeness of coverage changes over time, thus affecting the comparability and accuracy of the resulting estimates.

Farm registers based on administrative sources such as business registrations or tax collections. This process is used in some developed countries and offers advantages over the registers from the agricultural census, but again, they need to be updated regularly. A disadvantage of the administrative sources is that they may not include the total population, especially units below a threshold required to be registered or pay taxes. In other words, while they will be inclusive of commercial farms, they are not likely to include small-scale farms and subsistence farming units.

Area sample frames. An area sample frame is the land mass of the country or the space within a country containing the populations of interest and is suitable for obtaining information about variables associated with land such as crops, livestock, forests and water. Both maps and satellite images are used to divide the country into administrative areas such as provinces, districts, etc. Satellite imagery can also be used to subdivide the administrative areas into land-use categories, such as cropland, rangeland, woodlands, urban areas, etc. Sampling units of segments of land with identifiable boundaries can be formed, or each land-use stratum can be divided into square grids with a sample of points becoming the sampling units. During the data collection process, rules of association are used to connect farm holdings or households to the segments or points.

Depending on the process used, area frames can be costly and time consuming to construct. However, use of satellite imagery and two-stage sampling of points have reduced the cost and time. An added advantage of an area frame is that the frame does not go out-of-date; it is complete and provides a basis to georeference survey data with the underlying land use. It also provides ground truth useful for classifying satellite imagery by land cover. The primary disadvantage of area frames is that the sampling is based on land use and not on the size and type of agricultural holding.

Participatory data: helping communities to build datasets

Participatory 3D modelling (P3DM)[10] is a community-based tool – better defined as a 'process' – that integrates local spatial knowledge with data on land elevation and sea depth to produce a physical model. Local communities participate in the model building of the territory that they rely on for their livelihoods and cultural practices. A practice initially used in the late 1980s in Thailand to demonstrate where the Royal Forestry Department was developing plantations for catchment rehabilitation, it is geared towards community empowerment through measured, demand-driven, user friendly and integrated applications of geospatial technologies.

P3DM has already been used in a variety of contexts, to claim land ownership, to transfer knowledge across generations, and to manage conflicts. The approach has recently started to take off in the Pacific region, enabling people in small island countries – where rising sea levels

9 Adapted from World Bank (2011b).
10 This paragraph is adapted from Rambaldi (2015).

are posing a serious risk to many people's livelihoods – to make informed decisions about how best to manage risk and adapt to climate change.

In order to protect P3DM data, in 2006 the community of practitioners focusing on participatory GIS practice developed a guideline on practical ethics for participatory GIS (PGIS) practitioners, facilitators, technology intermediaries and researchers to stimulate the adoption of good practice. The guideline has been published in 12 languages and governs the way people undertaking participatory mapping should behave in the process of generating, handling, storing and sharing data. The code recommends that knowledge holders remain in full control throughout the process and that data are gathered and eventually shared with their free prior informed consent.

One of the most important components of a P3DM process is to involve external agencies from the very beginning, which can raise awareness among 'outsiders' about the depth, accuracy and relevance of local knowledge. This may induce a new sense of esteem for local knowledge holders.

The Ethiopian Commodity Exchange

In Ethiopia, small-scale farmers who produce 95% of Ethiopia's output generally receive little information regarding price and quality standards. Consequently, they often received prices far below market value – estimated on average to be 35–38% of export price – and also suffered serious losses when traders reneged on payment.

The Ethiopian Commodity Exchange (ECX) provides real-time price data to small-scale farmers with dissemination mechanisms tailored to farmer's needs allowing them to improve their decision-making. The ECX was set up in 2008 with an initial capital investment of US$20 million from the Ethiopian government, as well as from bilateral and multilateral donors.

Information is accessible via SMS text messages, an interactive telephone hotline, tickers using electronic displays at specific regional sites, the EXC.com website, as well as dissemination on TV, radio and newspapers in four languages.

In 2012, 70% of the 1.2 million monthly calls to the ECX data server came from rural areas. Improved knowledge about coffee prices reduced trader margins by almost half, with increased revenue going to farmers. In addition, the ECX created transparent quality standards, allowing farmers to increase product quality and profits. As a result, in the first year and a half of coffee trading, the Exchange saw the volume of highest grade coffee triple.

Currently the ECX plans to broaden the range of crops it trades (from coffee and sesame seeds to sugar and grains) and wants to introduce stocks and bonds under a five-year expansion plan.

Source: Adapted from Minney (2012).

CHAPTER 3

Conclusion

Issues and challenges around data

Data quality

Inaccurate or poor quality data can be misleading. The entire process of data design, collection, analysis and dissemination needs to be of high quality and integrity. A robust framework for quality assurance is required, particularly for official data, including internal systems as well as periodic audits by professional and independent third parties. Existing tools for improving the quality of statistical data should be used and strengthened, and data should be classified using commonly agreed criteria and quality benchmarks.

Governance

Although the majority of publicly available online data can potentially be used for development purposes, private-sector companies may be reluctant to share data due to concerns about competitiveness and customer privacy. It is therefore critical to establish and ensure a **legal framework** that defines rules for privacy-preserving analysis and to protect the competitiveness of private-sector companies willing to share data. Suitable legal frameworks, ethical guidelines and technological solutions for protected data sharing are central to leveraging big data for development. There are large amounts of data generated by individuals which are of great importance for collective use but mechanisms must be developed to ensure adequate user privacy and security.

Business models should also be created to provide the appropriate incentives for private-sector actors to share and use data for the benefit of society. Such models already exist in the internet environment. Companies offering search services (e.g. Google) and social networking (e.g. Facebook) at no charge to end users make money from the products they offer because the usage data these products generate is valuable to other ecosystem actors. Similar models could be created in the mobile data sphere, and the data generated through them could maximise the impact of scarce public sector resources by indicating where resources are most needed.

Protection and privacy

As more data becomes available in disaggregated forms, privacy issues are increasingly a concern about what data are collected and how it is used. There is a need to guarantee individuals' rights to control what information related to them may be disclosed, which has implications for all areas, from data acquisition and storage to retention, use and presentation. In many cases, the production of data itself raises concerns as people may be unaware of the data they are generating on a daily basis and not understand how it may be used. Clear international norms and robust national policy and legal frameworks need to be developed that regulate opt-in and opt-out, data mining, use, and data reuse for other purposes, transfer and dissemination. They should enable citizens to better understand and control their own data, and protect data producers from demands of governments and attacks by hackers, while still allowing for rich innovation in reuse of data for the public good.

Data disaggregation

As far as possible, and with due safeguards for individual privacy and data quality, data should be disaggregated across geography, wealth, disability, sex and age etc. based on relevance to the programme, policy or other matter under consideration. Disaggregated data can provide a better comparative picture of what works, and help inform and promote evidence-based policy-making at every level.

Cyber security risks

The risk of data breaches rises with the increase in volume and value of stored data. According to company surveys, reported thefts of electronic data surpassed physical property losses as the major crime problem for global companies for the first time in 2010 (Kroll, 2010). These are typical cyber espionage incidents often targeting a sector's key organisations or key competitors

to steal data or different forms of intellectual property and to reduce these organisations' competitive advantage. As data usage today requires information systems and networks to be more open, organisations are obliged to adapt their security policies to the more open and dynamic environment in which data are widely exchanged and used.

Digital divide and capacity development

Major gaps are already opening up between the data haves and have-nots. Without action, a whole new inequality frontier will open up, splitting the world between those who know, and those who do not. Many people are excluded from the new world of data and information by language, poverty, lack of education, lack of technology infrastructure, remoteness or prejudice and discrimination. According to McKinsey, African countries spend about 1.1% of GDP on investment in and use of internet services, less than a third, on average, of what is spent by richer countries – meaning that the gap in internet availability and use is growing every year.

Data rights

Human rights cut across many issues related to the data revolution. These rights include, but are not limited to the right to be counted, to privacy and to ownership of personal data, to participation, etc. Any legal or regulatory mechanisms, or networks or partnerships, set up to mobilise the data revolution for sustainable development should have the protection of human rights as a core part of their activities, specify who is responsible for upholding those rights, and should support the protection, respect and fulfilment of human rights (Data Revolution Group, 2014).

Timing

The value of data produced can be enhanced by ensuring there is a steady flow of high-quality and timely data from national, international, private big data sources, and digital data generated by people. Standards should be tightened and technology leveraged to reduce the time between the design of data collection and the publication of data. The data cycle must match the decision cycle.

Open data in the Caribbean

Historically, agriculture has played an important role economically and socially in the Caribbean, and many Caribbean countries have demonstrated a strong institutionalised commitment to freedom of information (FOI). As of 2011, of the 16 member countries of the Caribbean Community (CARICOM), 7 had enacted FOI laws, four had drafted FOI legislation, and two had guaranteed FOI as a constitutional right (Taylor, 2011). While this has not translated into the enactment of open data policies or representation in global initiatives such as the OGP (to date, only the Dominican Republic and Trinidad and Tobago have taken measures to join), sub-national institutional collaborations continue to gain momentum.

Across the region, members of the ICT, NGO and academic communities, in collaboration with government agencies, have begun to implement activities that illustrate the potential impact of 'openness' in governance, service delivery and public sector data on the key development challenges faced in the region. The Caribbean Open Institute (COI), a coalition of organisations supporting open approaches to addressing development issues in the region, was founded in 2011 with the support of the Canadian International Development Research Centre.

One of the largest COI initiatives has been the organisation of the Developing the Caribbean Open Data Conference and Code Sprint (DevCA), an annual regional open data and development event held simultaneously across multiple Caribbean islands. DevCA combines a two-day conference highlighting global best practices in the open data space and progress made regionally, with a 24-hour code sprint in which technologists and domain experts from across the region work on regional problems in thematic areas, such as tourism and agriculture, using Caribbean data.

Agricultural open data development in the Caribbean

While adoption of open data in the Caribbean remains nascent, momentum has been steadily building among key stakeholders in the region. Agriculture, in particular, has been the focus of many early initiatives.

In Trinidad and Tobago, the Caribbean ICT Research Programme at the University of the West Indies, St Augustine has made significant progress in the application of ICTs and open data within the Fisheries sector through its mFisheries project. mFisheries is a suite of mobile phone applications designed to

improve market conditions, efficiency and safety for small-scale fisherfolk in Trinidad and Tobago.

In Jamaica, an ongoing collaboration between the Mona School of Business and Management (MSBM), the SlashRoots Foundation and the Rural Agricultural Development Authority has focused on the issues of praedial larceny and access to information among agricultural stakeholders. The collaboration has resulted in the creation of HarvestAPI, an open data platform for publishing and sharing agricultural sector information, and Clip, an SMS information service that provides access to agricultural information via mobile phones.

More recently, a study of open data's value to the Jamaican economy was executed by the Caribbean Policy Research Institute and MSBM. The study evaluated the potential economic value of open data to the Jamaican economy to range from a lower bound of JAD 0.21 billion to an upper bound of JAD 2.42 billion[11]. Agriculture, a key industry in Jamaica, was selected as one of three sector analyses. The study estimated that open data would enable a 10% improvement in value-add from productivity gains in the agriculture sector alone – an additional contribution of approximately USD 92 million to Jamaica's GDP.

CTA'S AgriHack programme, baseline study on open data, and potential engagement with regional partners around open data are all very timely. These initiatives will capitalise upon the growing interest in both agriculture and open data. The research methodology combined an online survey with in-depth interviews with key institutional stakeholders. The online survey was distributed among a wide range of regional agricultural stakeholders and resulted in participation from 74 respondents.

Among agriculture stakeholders involved in both the online survey and in-depth interviews, there was a common consensus on the importance of data and information within the agriculture sector, and recognition of benefits open agricultural data benefits. Over 90% of respondents that answered the question indicated that they believed open data was 'very important' or 'important' to the agricultural sector within their country.

Survey respondents provided a variety of answers in support of its importance. Some of the perceived benefits of open data respondents identified include: (i) greater access to information for academic research policy or entrepreneurship; (ii) catalyst for economic activity and entrepreneurship; and (iii) increased farmer productivity. Similarly, institutional stakeholders often provided high-level responses to the perceived benefit of open data but seldom identified their own organisations as providing open data. However, when discussing how their organisations shared and utilised information, particularly with other government and research bodies, respondents were often more specific and illustrative in their responses. For instance, respondents from the Jamaican Meteorological Office and the Jamaican and Trinidadian Agricultural Information Systems provided examples of how public or private-sector organisations could make data requests, such as through weather analysis reports. This could indicate that while organisations are accustomed to sharing information among themselves and understand its value, they have not yet made the connection between this ongoing practice and opening data more broadly for consumption by innovators or the general public. While support for open data was high, only 17 of 46 survey respondents indicated that they knew of any publicly available agricultural datasets in the Caribbean (27 participants opted not to answer the survey question).

Data usability

Data are often presented in ways that cannot be understood by most people. The data architecture should therefore place great emphasis on user-centred design and user friendly interfaces. Communities of 'information intermediaries' should be fostered to develop new tools that can translate raw data into information for a broader constituency of non-technical potential users and enable citizens and other data users to provide feedback.

The way forward

As the volume, variety and velocity of data continues to increase, and the movement for open data gains momentum across different sectors, the debate is about how to unlock the untapped potential to innovate, experiment and adapt through the world of open data. Who really benefits from the opening of data? How can open data be applied to tackle global challenges such as agriculture and nutritional security? How to ensure that

11 Centre of Excellence for IT-enabled Innovation. 2014. *Open Government Data: A Catalyst for Jamaica's Growth and Innovation Agenda*. Kingston, Jamaica: Caribbean Policy Research Institute.

people, organisations and even governments are not left behind or even excluded from this data revolution (Data Revolution Group, 2014)?

These are some of the questions when tackling open data. The benefits of open data are expected to increase and as has been highlighted in this report, more organisations are adopting open data policies.

To achieve these benefits, a deeper understanding of several interdependent issues is needed on:

- The dynamics of constantly changing data flows
- New means and mechanisms for engaging individuals which are meaningful, pragmatic, adaptive and proportional
- The responsibilities of those who own and use data not only to develop new data tools and practices, but also to connect with other actors that can translate data into practical solutions and business models for improving agriculture and nutrition.

With so much uncertainty, the need for continuous experimentation, learning and sharing is paramount. Investing in small-scale pilots that bring together private sector, regulators, civil society and local communities will provide the insights and local knowledge critical for long-term resilience and adaptation (World Economic Forum, 2015). Efforts to improve cooperation between old and new data producers, ensure the engagement of data users, and develop global ethical, legal and statistical standards to improve data quality and protect people from abuses in a rapidly changing data ecosystem are needed. Brynjolfsson *et al.* (2011) estimate that the output and productivity of firms that adopt data-driven decision-making are 5% to 6% higher than would be expected from their other investments in and use of information technology. These firms also perform better in terms of asset utilisation, return on equity and market value. Growing investments in data management and analytics partly reflect the increasing economic role of data.

When referring to the **agriculture sector**, open data combined with agricultural knowledge, remote sensing, and mapping can support advice and early warnings for farmers (Syngenta Foundation, 2011). That information can be critical to protecting crops from pests and extreme weather, increasing yields, monitoring water supplies and anticipating changes brought on by climate change.

The idea of providing agricultural information freely is not entirely new. Data have been made available in the past by various agencies and research organisations, but the amount of data being generated now is increasing, as is our ability to share it effectively. This is leading to a growing interest in making good use of it (van Vark, 2013).

Farmers have traditionally been provided new information through linear, top down extension processes. In most developing counties it is the government and public sector that takes the main responsibility for providing information as a public service.

Source: CTA (2016)

Figure 4. Data needs of farmers.

However, the shift to market-oriented agriculture and the information revolution driven by new ICTs has triggered the emergence of pluralistic, networked extension and rural advisory services that use many means for farmers, actors and other agricultural stakeholders to access, share and exchange information (Maru, 2014).

The rights of farmers to data, information and knowledge have many dimensions to be considered. Sharing data more openly has the potential to bring better services for farming such as in the access and use of farm inputs, in managing production and in marketing produce. The key issues for farmers related to data, information and knowledge include the need for it to be available, accessible in a timely, fair and equitable manner, and to be affordable, relevant, useful and trustworthy for famers to effectively use it. Farmers also need to have the capacity to use the collated data, information and knowledge. They should also be included in processes related to deciding on the information, data and knowledge to be generated, shared and exchanged, according to their needs and preferences.

Yet, farmers are not only providers and recipients of data. They are also providers of traditional knowledge, innovations and practices relevant for food production and agriculture. Farmers have the right to give their prior informed consent for the access and use of their knowledge, innovations and practices, as well as to participate in a fair and equitable way in the benefits arising from the use of their knowledge, innovations and practices. Therefore, farmers need to be fully involved in the process of management and use of data. This process can actually be put in place by enhancing farmers' capacities to access, understand and manage data, e.g. through the use of ICT tools, the development of standards, integration and exchange of data and information.

For researchers and governments seeking to use open data to improve agriculture, food and nutrition, the integration of the open data approach will largely depend on the quality of the data and the cost-effectiveness of turning it into a usable format. It is important to note that more data does not automatically translate into better decisions. Collecting and processing data and turning them into information, using data and making them open for others to use and reuse is an indispensable investment to get information for actionable data by policy-makers.

Therefore, governments should empower and finance independent statistical offices to address the opportunities derived from the use of open data and innovation and to adapt to the new world of data to collect, process, disseminate and use high-quality, open, disaggregated and geo-coded data, both quantitative and qualitative (Open Data for Development, 2015).

Governments should also design and implement governance systems which regulate the use and management of data. Development partners should also support data investment, especially for capacity building, and infrastructure development in order to get qualitative data usable to demonstrate their impact. Governments should also support the setting and enforcement of common standards for data collection, production, anonymisation, sharing and use to ensure that new data flows are safely and ethically transformed into global public goods, and maintain a system of quality control and audit for all systems and all data producers (Data Revolution Group, 2014).

As it stands, the benefits of open data are largely accrued to those with the resources to access data and the capacities to analyse it. The concentration of benefits towards those with power, resources, and access to ICTs will be important considerations for applying open data in the agriculture and nutrition sectors. The realities of agrarian change, the mass exoduses out of agriculture in many countries, and the ownership of land and resources will influence who benefits from the application of open data.

The gendered agricultural practices and access and control over resources also mean that applications may need to specifically be designed and target youth and women smallholders.

References

Addison, C., Boto, I., Mofolo, L. 2015 *Data: the next revolution for agriculture in ACP countries.* Reader, Brussels Briefing. 40. Available at: https://brusselsbriefings.files.wordpress.com/2015/01/reader-bb40-data-revolution-eng_v2-low-res-proof2-2.pdf [Accessed February 2016].

AfDB (African Development Bank). 2013a. *Improving statistics for food security, sustainable agriculture and rural development*. ESTA Department, African Development Bank, Tunis, Tunisia. Available at: http://bit.ly/1U4Grns. [Accessed April 2016].

AfDB (African Development Bank). 2013b. *Situational Analysis of Economic Statistics in Africa: Special focus on GDP measurement*. African Development Bank, Tunis, Tunisia.

Brynjolfsson, E., Hitt, L. and Heekyung, K. 2011. *Strength in Numbers: How Does Data-Driven Decision-making Affect Firm Performance*. Available at: http://www.a51.nl/storage/pdf/SSRN_id1819486.pdf [Accessed April 2016].

Carletto, C., Jolliffe, D. and Banerjee R. 2015. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *In* Jerven, M. and Johnston, D. (eds). S*tatistical Tragedy in Africa? Evaluating the Data Base for African Economic Development*, Special Issue, *Journal of Development Studies*.

Carletto, G., Himelein, K., Kilic, T., Murray, S., Oseni, M., Scott, K., and Steele, D. 2010. *The Living Standards Measurement Study – Integrated Surveys on Agriculture*. Available at: http://www.fao.org/fileadmin/templates/ess/pages/rural/wye_city_group/2010/May/WYE_2010.2.1_Carletto.pdf [Accessed on October 2015]

CIARD. 2014. *CIARD Manifesto. Towards a Knowledge Commons on Agricultural Research for Development*. Available at: http://www.ciard.net/about/manifesto [Accessed April 2016].

CTA. 2016. *Open data for agriculture and nutrition*. Available at: http://www.cta.int/en/article/2016-03-03/open-data-for-agriculture-and-nutrition.html [Accessed April 2016].

Data Revolution Group. 2014. *A world that counts*. Available at: http://www.undatarevolution.org/report/ [Accessed on April 2016].

Davies, T. (ed.). 2013. *Open Data Barometer. 2013 Global Report.* http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf [Accessed November 2014].

Duflo, E., Kremer M. and Robinson, J. 2008. *How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya* American Economic Review, 98 (2): 482–8

European Commission. 2011. *Open Data. An Engine for Innovation, Growth and Transparent Governance.* Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF [Accessed April 2016].

FAO.2006. *Independent External Evaluation of the FAO, Inception Report submitted to the Council Committee for the Independent External Evaluation (CC-IEE) by the Independent External Evaluation Core Team*. Food and Agricultural Organization, Rome, Italy

FAO. 2009. *2050: A third more mouths to feed*. Available at: http://www.fao.org/news/story/en/item/35571/icode/ [Accessed April 2016].

FAO and AfDB (African Development Bank). 2012. *Improving Statistics for Food Security, Sustainable Agriculture, and Rural Development. An Action Plan for Africa 2011–2015.* Available at: http://www.fao.org/docrep/016/am084e/am084e.pdf [Accessed April 2016].

FAO 2016. *FAO and Google are working together to make high-resolution satellite data an everyday tool in managing the world's natural resources*. Available at http://www.fao.org/news/story/en/item/410307/icode/ [Accessed April 2016].

Fonzi, C.J. and Chau, V. 2014. *Smallholder Impact and Risk Metrics: A Labyrinth of Opportunity.* Briefing 03. The Initiative for Smallholder Finance. Available at: http://bit.ly/1SuShWW [Accessed April 2016].

Jellema, A., Meijninger, W. and Addison, C. 2015. *Open Data and Smallholder Food and Nutritional Security*. CTA Working Paper. Available at: http://www.cta.int/images/Opendataforsmallholders-report_.pdf [Accessed on February 2016].

Jerven, M. 2014. 'The Political Economy of Agricultural Statistics and Input Subsidies: Evidence from India, Nigeria and Malawi.' *Journal of Agrarian Change*, 14(1): 129–145.

Jerven, M. and Ebo Duncan, M. 2012. 'Revising GDP estimates in sub-Saharan Africa: Lessons from Ghana.' *African Statistical Journal*, 15: 13–22. Available at: http://bit.ly/1Nkm7Nm [Accessed April 2016].

Kondylis, F. 2015. *Measuring Yields from Space*. Published on *Impact Evaluations*. Available at: http://blogs.worldbank.org/impactevaluations/publicsphere/endpovertyinsouthasia/publicsphere/measuring-yields-space [Accessed April 2016].

Kroll. 2010. *Global Fraud Report. 2010-2011*. Available at: http://fr.kroll.com/media/pdfs/FraudReport_English-US_Oct10.pdf [Accessed April 2016].

Maru, A. 2014. *Rights of Farmers for Data, Information and Knowledge. A CIARD E-Discussion.* Available at: http://www.ciard.info/news-and-events/blog/rights-farmers-data-information-and-knowledge-ciard-e-discussion [Accessed April 2016].

McNaughton, M. and Soutar, D. 2015. *Agricultural Open Data in the Caribbean*. CTA Working Paper 15/02. Available at: http://www.cta.int/images/Open_Data_Carib_new_cover.pdf [Accessed April 2016].

Minney, T. 2012. 'Making the market, the story of the Ethiopian Commodity Exchange.' *African Capital Markets News* [online]. Available at: http://www.africancapitalmarketsnews.com/1576/making-the-market-the-story-of-the-ethiopian-commodities-exchange/ [Accessed April 2016].)

Mutuku, L. and Mahihu, C.M. 2014. *Open Data in Developing Countries: Understanding the Impacts of Kenya's Open Data Applications and Services.* Available at: http://www.opendataresearch.org/content/2014/731/understanding-impacts-kenya-open-data-applications-and-services [Accessed November 2015].

Ologunagba, C. 2015. *Tracking progress under the MDGs hampered by data gaps – Official*. News Agency of Nigeria. Available at: http://nannewsnigeria.com/tracking-progress-under-mdgs-hampered-data-gaps-official [Accessed April 2016].

Open Data for Development. 2015. *Enabling the Data Revolution. An international Open Data Roadmap*. Conference Report. 3rd International Open Data Conference, Ottawa, Canada, May 2015. Available at: http://opendatacon.org/report/ [Accessed April 2016].

Open Knowledge Foundation. 2012. *The Open Data Handbook.* Available at: http://opendatahandbook.org/ [Accessed November 2015].

Open Source. 2014. *What is open source?* Available at: http://opensource.com/resources/what-open-source [Accessed April 2016].

Precision Agriculture. 2014. *Review: Crop Monitoring Systems.* Available at: http://precisionagricultu.re/review-crop-monitoring-systems/ [Accessed April 2016].

Rambaldi, G. 2015. 'Participatory Data.' *ICTUpdate* 79. Available at: http://ictupdate.cta.int/en/Feature-Articles/Participatory-data [Accessed April 2016].

SDSN (Sustainable Development Solutions Network). 2015. *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development*. Available at: http://unsdsn.org/wp-content/uploads/2015/04/Data-for-Development-Full-Report.pdf. [Accessed April 2016].

Syngenta Foundation. 2011. *Mobile Application in Agriculture*. Available at: http://www.syngentafoundation.org/__temp/Report_on_mAgriculture_abridged_web_version.pdf [Accessed April 2016].

Tatem, A.J., Huang, Z., Narib, C., Kumar, U., Kandula, D., Pindolia, D.K., Smith, D.L., Cohen, J.M., Graupe, B., Uusiku, P. and Lourenço, C. 2014. 'Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning.'

Malaria Journal, 13: 52. Available at: https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-13-52 [Accessed April 2016].

Taylor, K.C. 2011. *A Review of Freedom of Information, Data Protection and Open Data in the Caribbean*. Lecture, Internet Governance Forum, Nairobi, Kenya.

United Nations Global Pulse. 2014a. *Global Pulse Annual Report 2014.* Available at: http://www.unglobalpulse.org/2014-Annual-Report [Accessed February 2016].

United Nations Global Pulse. 2014b. *Responsible Data Forum on Private Sector Data Sharing – Event Summary.* Available at: http://www.unglobalpulse.org/RDF-private-sector-data-summary/ [Accessed April 2016].

van Vark, C. 2013. *How Might Open Data in Agriculture Help Achieve Food Security?* Available at: http://www.theguardian.com/global-development-professionals-network/2013/nov/25/open-data-food-security-agriculture [Accessed April 2016].

Venkatalakshmi, B. and Devi, P. 2014. 'Decision support system for precision agriculture.' *International Journal of Research in Engineering and Technology,* 3(7): 849–852. Available at: http://www.academia.edu/8160091/DECISION_SUPPORT_SYSTEM_FOR_PRECISION_AGRICULTURE [Accessed April 2016].

Wall, M. 2014. *Big Data: Are you ready for blast off?* BBC News [online]. Available at: http://www.bbc.com/news/business-26383058 [Accessed April 2016].

World Bank. 2011a. *ICT in Agriculture: Connecting Smallholders to Knowledge, Networks, and Institutions*. e-Sourcebook. Available at: http://www.ictinagriculture.org/sites/ictinagriculture.org/files/final_book_ict_agriculture.pdf [Accessed April 2016].

World Bank. 2011b. *Global Strategy to Improve Agricultural and Rural Statistics*. Report No. 56719-GLB. Economic and Sector Work. World Bank, Washington, DC. Available at: http://bit.ly/26jn9Q9 [Accessed April 2016].

World Bank. 2014. *Open Data in 60 Seconds*. Available at http://toolkit.dev.zognet.net/en/open-data-in-60-seconds.html [Accessed April 2016].

World Economic Forum. 2015. *Data-Driven Development Pathways for Progress*. Available at: http://www3.weforum.org/docs/WEFUSA_DataDrivenDevelopment_Report2015.pdf [Accessed April 2016].

ANNEX 1

Glossary: A selection of key concepts[12]

Aggregate data

A combination of unit records created with the objective that individual details are not disclosed.

Big data

A loose term, not formally defined, for high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing, that can give enhanced insight and decision-making.

Crowdsourcing[13]

The practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers.

Data

The quantities, characters or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical or mechanical recording media. (The terms data, information and knowledge are frequently used for overlapping concepts. The main difference is in the level of abstraction being considered. Data is a broad term, embracing others, but is often the lowest level of abstraction, information is the next level and, finally, knowledge is the highest level.) See also Raw data.

Dataset

A collection of data, usually presented in tabular form, presented either electronically or in other formats.

Generic Statistical Business Process Model[14]

The Generic Statistical Business Process Model is a means to describe statistics production in a general and process-oriented way. It is used both within and between statistical offices as a common basis for work with statistics production in different ways, such as quality, efficiency, standardisation and process-orientation. It is used for all types of surveys, and 'business' is not related to 'business statistics' but refers to the statistical office, simply expressed.

Open data

Data are open if anyone is free to access, use, modify and share it – subject, at most, to measures that preserve provenance and openness.

Public domain

Works that are publicly available and in which the intellectual property rights have expired or been waived.

Open access (academic)

Provision of free access to peer-reviewed academic publications.

Open Data Commons

An Open Knowledge Foundation project run by its Advisory Council and like the Foundation is a not-for-profit effort working for the benefit of the general open knowledge community. Open Data Commons is the home of a set of legal 'tools' to help others provide and use open data.

Open Government Licence

The Open Government Licence offers a legal solution to enable the provision and use of public-sector information under a common set of terms and conditions.
It enables any public-sector information-holder to make their information available for use and reuse under its terms. The main

12 These concepts are drawn from the Open Data Glossary of the United Kingdom Government (https://data.gov.uk/glossary) unless otherwise specified.
13 http://www.merriam-webster.com/dictionary/crowdsourcing
14 https://ec.europa.eu/eurostat/cros/content/gsbpm-generic-statistical-business-process-model-theme_en

requirement for re-users is to attribute the information provider and source.

Open source[15]

Within the software community 'open source' denotes software made available under a licence that enables anyone to access, modify and use its underlying code

Personal data

Data which relate to a living individual who can be identified – (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

Raw data

Raw data are data that have not been subjected to processing or any other manipulation beyond that necessary for its first use. Raw data, i.e. unprocessed data, is a relative term; data processing commonly occurs by stages, and the 'processed data' from one stage may be considered the 'raw data' of the next.

15 GODAN. Open Data Institute. 2015.How can we improve agriculture, food and nutrition with open data? Online at http://www.godan.info/wp-content/uploads/2015/04/ODI-GODAN-paper-27-05-20152.pdf

The Technical Centre for Agricultural and Rural Cooperation (CTA) is a joint international institution of the African, Caribbean and Pacific (ACP) Group of States and the European Union (EU). Its mission is to advance food and nutritional security, increase prosperity and encourage sound natural resource management in ACP countries. It provides access to information and knowledge, facilitates policy dialogue and strengthens the capacity of agricultural and rural development institutions and communities. CTA operates under the framework of the Cotonou Agreement and is funded by the EU.

FOR MORE INFORMATION ON CTA, VISIT WWW.CTA.INT

CONTACT US
CTA
PO Box 380
6700 AJ Wageningen
The Netherlands

Tel: +31 317 467100
Fax: +31 317 460067
Email: cta@cta.int

www.facebook.com/CTApage
@CTAflash