Module 15 USING ICT FOR REMOTE SENSING, CROWDSOURCING, AND BIG DATA TO UNLOCK THE POTENTIAL OF AGRICULTURAL DATA

JOSH WOODARD (FHI 360), with contributions from MECHTELD ANDRIESSEN (eLEAF), COURTNEY COHEN (aWhere), CINDY COX (IFPRI), STEFFEN FRITZ (IIASA), DREW JOHNSON (TechnoServe), JAWOO KOO (IFPRI), MORVEN MCLEAN (ILSI Research Foundation), LINDA SEE (IIASA) TARAH SPECK (aWhere), and TOBIAS STURN (IIASA).

OVERVIEW

By 2050, the global population is projected to reach approximately 9 billion. Population growth will be concentrated in poorer countries, particularly the low-income countries of Sub-Saharan Africa. By some estimates, agricultural productivity will need to double to meet everyone's needs for food (Foley 2014). For instance, if current trends continue, yields of the world's foremost food crops—maize, rice, wheat, and soybeans, which supply roughly two-thirds of calories consumed globally—appear likely to grow significantly more slowly than required to meet the projected global demand in 2050 (Ray et al. 2013). Some productivity growth will come from using more of the world's arable land for agriculture, but most of the available arable land is unevenly distributed, and about half of it is found in only seven countries.

If agricultural productivity is to grow sufficiently to meet the world's expanding demand for food, producers must be able to increase yields and cropping intensity (Alexandratos et al. 2012), improve the productivity of their livestock, and quite possibly diversify their portfolio of economic activities on and off of the farm. At the same time, producers are only too aware of the challenges and risks presented by a changing climate and growing population pressure. Multiple approaches are needed to support their efforts, including those to improve natural resource and farm management, to develop better crop varieties and animal breeds, to devise and use innovations in crop and livestock production (such as technologies for precision agriculture or livestock identification and tracking), to generate and share knowledge, and to improve access to markets, among others.

This module was originally written in 2015, with minor revisions in early 2016. All content was accurate at the time of writing.

The reviewers of this module were Lyudmila Bujoreanu (World Bank Group), Terhi Elisa Havimo (World Bank Group), Holger Kray (World Bank Group), Trevor Monroe (World Bank Group), and Eija Pehu (World Bank Group).

Virtually all of these paths to offsetting current trends and increasing the growth of agricultural productivity can benefit from revolutionary changes in how data are collected, generated, shared, analyzed, and visualized. Agriculture is notoriously complex, characterized by wide variation across time and space in terms of producers, production systems, biophysical conditions, and myriad other variables, at scales smaller than the smallest plot or larger than multiple agroecologies. Advances in information and communication technology (ICT) over the past 20 years have enabled individuals to gather, analyze, and share data more effectively, as well as to visualize and understand, as never before, what this information means for agriculture. The capacity to capture and analyze data has been growing exponentially with the global spread of relevant ICT tools, including geospatial statistical methods.

Big data (the proliferating types and amounts of data being collected), together with advanced ICT capabilities (such as more sophisticated computer processors and algorithms), are providing a more accurate understanding of existing conditions and generating better predictions of future conditions, enabling more informed (often real-time) decision making. In agriculture, ICT and big data are helping to leverage the global engagement of development practitioners, researchers, scientists, and producers across borders. New approaches combining enhanced productivity with environmental sustainability are being developed.

For example, since the 1990s, commercial producers in high- and middle-income countries have increasingly taken advantage of precision farming technologies such as global positioning systems (GPS), geographic information systems (GIS), remote sensors, and satellite imagery to improve productivity (box 15.1). Such technologies are increasingly within reach in low-income economies. In 2005, the median price of a computer was US$1,500 and a GPS device cost more than US$2,000 (Martin et al. 2005); a decade later, a smartphone with a GPS receiver and more computing power than a computer available in 2005 cost less than US$100. The array of sensors in smartphones has expanded to include barometers and thermometers that can collect hyperlocalized weather information. Small-scale producers with access to mobile phones are beginning to benefit from improved tools also for networking, decision making, and analysis based on these technologies.

The same big data sets are used by a wide array of stakeholders in distinctly different ways. With hyperlocalized weather data, for instance:

- Farmers can make better planting decisions based on more accurate weather predictions and better prepare themselves to adapt to changing conditions.
- The Ministry of Agriculture can use more accurate information on local weather patterns and disease and pest populations to tailor the extension messages it provides to producers.
- Government agencies responsible for disaster risk reduction and response can use this information to better allocate resources based on hyperlocalized weather events.
- Donors can use this data to design programs that are better attuned to local needs.
- The private sector can use these data to offer a whole range of services that it could not deliver previously, such as weather insurance for smallholder farmers. Two examples are the Agriculture and Climate Risk Enterprise Ltd. (ACRE) in East Africa and the weather-based crop insurance offered by private insurance providers in India (Greatrex et al. 2015).[1]

Big data analytics makes it increasingly possible to combine multiple types of data in a single interface, often referred to as a mashup, which improves the prospects of gaining insights that would not have been accessible before. A growing community of scientists and agricultural development practitioners is able to perform more complex analyses using different types of data generated in very different ways and assembled in global data sets, which are increasingly available to the public (box 15.2). For example, analyses combining data from the World Bank's Living Standards Measurement Study–Integrated Surveys on Agriculture (LSMS–ISA) with satellite or aerial images could potentially help policy makers to develop a more robust picture of the challenges facing various types of producers than they could derive from either set of data alone.[2]

1 ACRE was previously known as Kilimo Salama.
2 To learn more about LSMS-ISA, see http://go.worldbank.org/BCLXW38HY0.

BOX 15.1. The Intersection of Remote Sensing, Crowdsourcing, and Big Data

In the United States, the agricultural firm John Deere exemplifies how remote sensing, crowdsourcing, and big data can be combined to offer farmers precise insights to increase efficiency and productivity. The company's online portal pulls in data from farmers' sensors. The aggregated data from thousands of farmers are combined with external data sets (such as weather data) and, powered by big data analytics, used to advise farmers on what to plant on certain areas of their land. The system can also predict when machinery is likely to break and notify a nearby parts distributor to stock that part. Despite the obvious benefits to farmers, some farmers have questioned whether they should be compensated for how John Deere uses their data to enhance its commercial services.[a] This question needs to be addressed as similar types of services and the business models that support them are rolled out throughout the world, especially in middle- and low-income countries where producers have far more limited potential to pay for subscription services.

Source: Authors.
a. Marr 2015.

BOX 15.2. Visualizing Data Sets for Development

The World Bank's Spatial Agent app enables users to access global data sets in the public domain and to display the data in a graphic or other pictorial format that makes the implications easier to understand. The Spatial Agent app includes spatial and temporal data from a number of global institutions and is available for free for Android and iOS mobile operating systems.

Source: Authors.

Both the public and private sectors clearly recognize the potential value of investing in big data aggregation and analysis for agriculture. The agricultural technology sector, which includes the ICT applications highlighted in this module, received more than US$2.3 billion in investments in 2014, surpassing investments in financial technology and clean technology in that same year. The level of investment in 2014 represented an increase of 170 percent over the previous year, and this strong growth trend appears likely to continue

BOX 15.3. Separating the Hype from Reality

It is easy to understand why some people are skeptical about the potential of remote sensing, crowdsourcing, and big data to improve the lives of the poorest producers. The reality is that even as global mobile phone penetration rates continue to grow, a significant number of poor producers, especially women, do not own a phone, much less any way of accessing the benefits of remote sensing and big data.

To some, these technologies represent a luxury reserved for high-income countries and affluent producers, rather than technologies with practical applications for less thriving economies and producers. There is some truth to that sentiment. The World Economic Forum's 2014 Networked Readiness Index identifies a number of countries, particularly in Sub-Saharan Africa, that are lagging in leveraging the potential of ICT.[a]

Because advances in some technologies tend to move faster than many can imagine, they may become more readily available throughout the world over the next few years. Smallholders may not actually use remote sensors and big data analytics themselves, yet it is very likely that—at the very least—the institutions that support those producers will be able to access such technologies and use them to benefit the individuals they serve.

Policy makers and donors who ignore these factors and simply write off the technologies mentioned in this module as a pipe dream do so at the risk of exacerbating the digital divide. It is important to examine how smart investments can leverage these technologies over the short and long terms to benefit the agricultural sector as a whole and especially poorer producers, many of whom have yet to benefit. This module should be seen as the beginning, not the end, of examining the prospects for applying these technologies, given that their evolution and deployment are so dynamic.

Source: Authors.
a. Bilbao-Osorio et al. 2014.

(Leclerc and Tilney 2015). Notable investments since 2013 have included:

- Monsanto's US$1 billion acquisition of the Climate Corporation, an agriculture analytics and crop insurance company that uses weather data (Tsotsis 2013).
- An investment of US$95 million in Planet Labs, which operates a legion of Earth-imaging microsatellites used by several sectors, including agriculture (Lawler 2015).
- Qualcomm Ventures' US$50 million investment into 3D Robotics, a drone manufacturer whose products are used, among other things, for precision agriculture (Burns 2015).
- Google Ventures' US$15 million investment in the Farmers Business Network, a massive farming database and decision-making tool (Reader 2015).
- The £12 million invested by the UK government in the Centre for Agricultural Informatics and Metrics of Sustainability, which will focus on using big data analytics to support agricultural development (Crawford et al. 2015).
- Uruguay's investment of US$55 million, with support from the World Bank, in the Sustainable Management of Natural Resources and Climate Change Project, which includes the development of a National Agricultural Information System decision-support tool.[3]
- The Government of India's launching of the Digital India initiative. Although not specific to agriculture, it includes a crowdsourcing platform (mygov.in) to gather citizens' feedback.

A strong case can also be made for public investments in big data as a public good. These kinds of innovative investments have a strong history of success that includes Landsat, the Global Agricultural Monitoring system, the Famine Early Warning Systems Network, and ALEXI—and that belies skepticism that remote sensing, crowdsourcing, and big data analytics can benefit low-income economies and the poorest producers (box 15.3). Given the very large investment required to support some types of remote sensors and big data services, it may be some time before commercially available service providers find that it pays to target small-scale producers in Sub-Saharan Africa and Asia. Deeper analysis may reveal particular opportunities for public investment to yield positive returns for society—for example, investments in services that are highly likely to improve smallholders' productivity but relatively unlikely to be commercially viable.

3 Learn more here: http://www.worldbank.org/projects/P124181/sustainable-management-natural-resources-climate-change?lang=en.

Key Challenges and Enablers

The world currently produces a lot of data, and this amount will continue to grow exponentially over the coming decades. In 2013, the world's stock of digital data was estimated to be around 4.4 zettabytes.[4] By 2020, that amount is expected to increase tenfold, to 44 zettabytes, with most of those data coming from emerging markets. Part of that growth is expected to come from data from embedded systems, as more and more devices begin communicating directly with each other.[5] This phenomenon—referred to as the Internet of Things (IoT)—enables devices to share data directly, without a human intermediary, although by 2020 the vast majority of data (90 percent) will still come from humans (IDC 2014).

Access to data also remains a challenge. Many governments and organizations, including the World Bank, have promoted open data, yet much of the world's data remains proprietary or exists in inaccessible formats.[6] Despite efforts to promote the opening of data for the public good, significant portions of the world's digital data are likely to stay outside the realm of public use for some time.

Agriculture is no stranger to this challenge. The collection and management of agricultural data is often fragmented among government agencies, development practitioners, and agribusinesses. Centralized and comprehensive agricultural databanks remain the exception rather than the rule. A recent commentary on data-driven agriculture in Nigeria notes that "the dearth of information is making it difficult to translate data into useful information for producers and other players in the value chain" (Essiet 2015).

While a reluctance to share data sometimes adds to this challenge, it is important to recognize that the problem stems largely from barriers related to data standards and the lack of interoperability. Hardware and software systems for collecting agriculture-related data are not all interoperable, meaning that they use incompatible formats. In some instances, standards that would facilitate interoperability between different systems are lacking as well. Public institutions can play a significant role, as they have already done, by promoting the use of open data sharing and standards. The Global Open Data for Agriculture and Nutrition (GODAN) initiative was launched in 2013 and now has over 100 partners from the public and private sectors. GODAN advocates for open data and open access policies in the public and private sectors, as well as for the release and reuse of data for social, economic, governance, and environmental benefit.[7] Another example is the Open Ag Data Alliance, launched in 2014 to help farmers access and control their data "by building an open source framework and a community of commercial vendors, farmers, academics, and developers upon which the emerging ag data market can rapidly grow." Some of the world's largest agricultural companies support the Open Ag Data Alliance, which can potentially serve as a model for the types of collaboration required to overcome the challenges of fragmented and unusable data.[8]

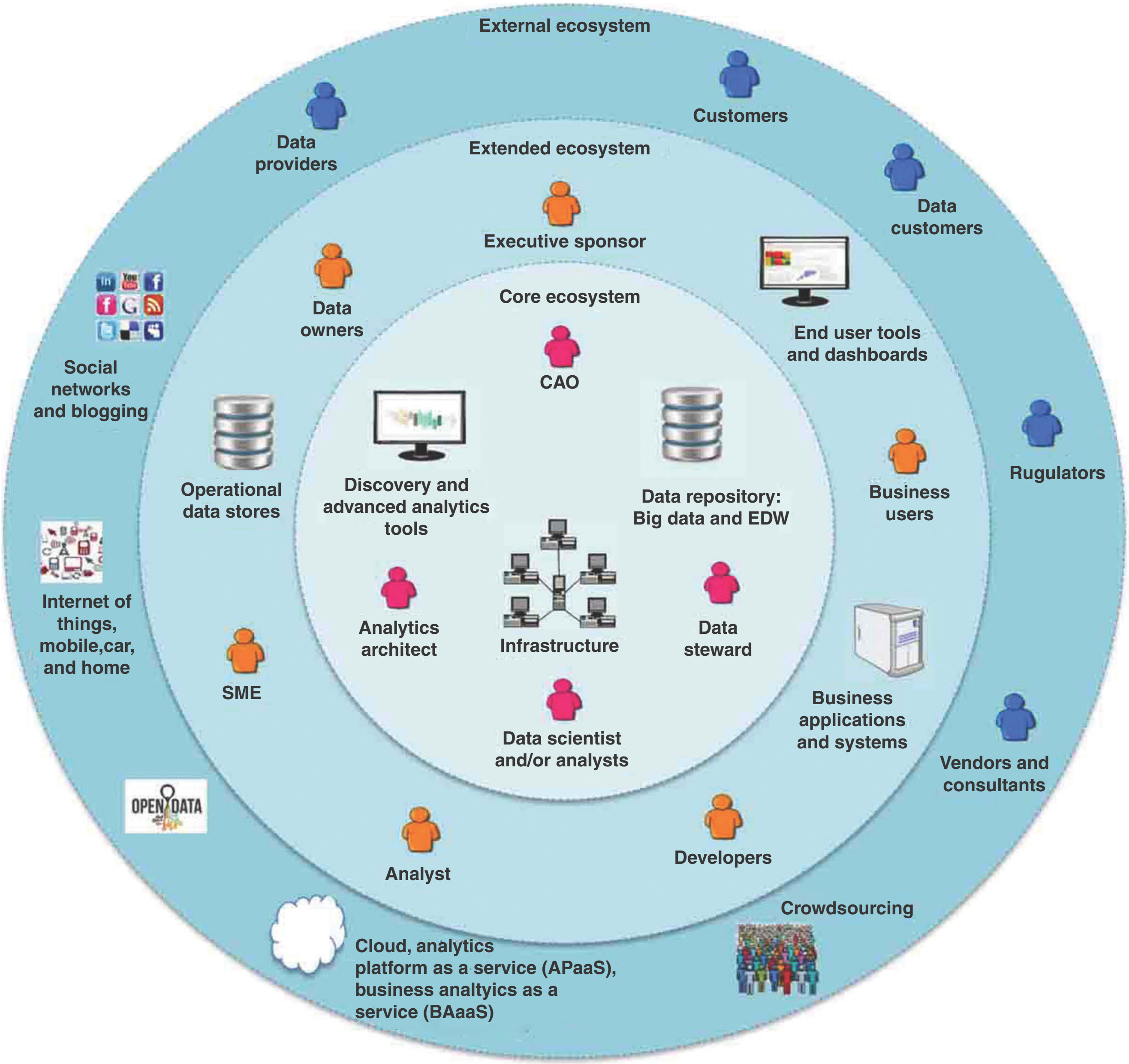
Aside from the technical barriers to using data, there are significant skill barriers. Many organizations find it challenging to hire individuals with the right experience to fully harness the value of data. The information technology research firm Gartner has estimated that 4.4 million jobs would be created in 2015 to support big data, but that only one-third of those positions would actually be filled due to limited talent within the industry (Gartner 2012).

Both the public and private sectors have important roles in helping to bridge these gaps in human and technological resources. Public and private educational institutions, with encouragement from governments where necessary, need to be preparing more students to take on the development of next-generation remote sensors and build a robust big data analytics ecosystem (figure 15.1 provides more detail on the components of a big data ecosystem). Organizations and companies will need to provide on-the-job training to ensure that their employees are equipped to use the new systems and processes introduced as a result of the trends highlighted in this module.

Informed consent and fair compensation for data collection are two other significant challenges. The issue of informed consent is particularly thorny when data are collected without a human intermediary—for example, through a crowdsourcing platform or via a short messaging service (SMS) survey in which participating individuals may not fully read the lengthy terms of service that convey this information. The development community is only just beginning to explore these issues seriously. In late 2014 in Nairobi, for instance, the Responsible Data Forum hosted an event in partnership with Amnesty International on "Consent and Crowdsourcing."

4 One zettabyte is the equivalent of just under 1 trillion gigabytes.
5 Embedded systems are computing systems within a device, such as a computing system within a refrigerator.
6 Open data are defined as data that are made freely available for use, reuse, and distribution, for any purpose, without restriction. The World Bank's Open Government Data Toolkit is a good starting point for learning more about open data in practice. It is available for free online at http://opendatatoolkit.worldbank.org/en/.
7 To learn more about GODAN, see http://www.godan.info/about/statement-of-purpose/.
8 To learn more about the Open Ag Data Alliance, see http://openag.io.

FIGURE 15.1. Example of a Big Data Analytics Ecosystem



Source: IBM Big Data & Analytics Hub.[10]
Note: Used with permission. Further permission required for reuse.

In examining that topic, participants attempted to address some of the following questions:[9]

- Do organizations that collect and use crowdsourced information have a responsibility to disclose those facts to users? If so, might disclosure affect the quality and quantity of information collected?
- Can ethics-based and consent requirements be built right into software, or do these responsibilities always lie with the organization using the software?
- Is consent more appropriate as a precondition for data collected for "research" versus other purposes?

Not all companies or organizations that offer data-enabled services to farmers are explicit about how they seek informed

9 For details on this event, see https://responsibledata.io/forums/consent-and-crowdsourcing/. The outputs of the event should eventually appear at https://wiki.responsibledata.io/RDF_Nairobi.

10 This graphic originally appeared in the article "Going Beyond Data Science toward an Analytics Ecosystem: Part 2" by Ahmed Fattah, which was published on the IBM Big Data & Analytics Hub Blog on March 14, 2014.

BOX 15.4. Monitoring and Evaluating Investments in Remote Sensing, Crowdsourcing, and Big Data for Analytics

The monitoring and evaluation of investments in any of the three big data trends covered in this module will vary, depending on the extent of the investment and anticipated outcomes. Even so, the following common indicators of outputs and outcomes are likely to be useful:

- **Remote sensing**
  - *Outputs:* Amount of data collected.
  - *Outcomes:* Impact of remote sensing on productivity and/or cost.
- **Crowdsourcing**
  - *Outputs:* Number of people engaged; number of contributions; amount of data collected.
  - *Outcomes:* Impact of contributions on overall outcome being tracked; cost-benefit comparison between crowdsourcing and traditional methods.
- **Big data for analytics**
  - *Outputs:* Number of insights gained; number of insights deemed accurate.
  - *Outcomes:* Impact of insights on efficiency or cost of overall outcome; changes in policy.

Source: Authors.

consent from farmers to use their data. Plans for using individuals' data might be contained in a service agreement that marginally literate farmers cannot read or understand. It behooves any development practitioner who promotes the use of ICT by farmers to understand exactly how farmers' data will be used and who owns their data, and in turn to clearly explain these issues to farmers so that they can give their truly informed consent.

The related question is what rights individuals have to receive some portion of the value generated from their data. For instance, a hypothetical weather service could generate millions of dollars in revenue from the sale of its hyperlocalized weather data. The data that were fed into the service's algorithms were crowdsourced from local farmers, who received free weather information in exchange for their participation. Should those farmers also have some right to compensation for the revenue generated by the service, or is the value of free weather information a fair exchange?

Resolving these and other questions related to informed consent, privacy, and ownership rights must be at the center of organizational policy discussions in relation to data collection from crowdsourcing and remote sensors.[11]

11 The book *Ways to Practise Responsible Development Data* is a great resource for learning more about responsible data practices. Download it for free at https://responsibledata.io/ways-to-practise-responsible-development-data/.

As the emphasis on using data and evidence on crafting development interventions is relatively recent but growing, we have pooled together a few key questions to support the monitoring and evaluation of data-centered approaches in box 15.4.

ORGANIZATION OF THIS MODULE

The topic notes that follow cover three ICT trends that are enabling producers, agribusinesses, policy makers, researchers, and agricultural development practitioners to optimize resources, connect people, and overcome data fragmentation: (1) remote sensing, (2) crowdsourcing and crowdmapping, and (3) big data for analytics. Although discussed separately, these trends are interrelated. The first two trends are driving the production of massive amounts of raw agricultural data, and big data analytics is the process through which these data is meshed, refined, and analyzed. Each topic note is followed by one or more innovative practice summaries highlighting a particular application or aspect of the topic at hand. Box 15.5 defines some key terms covered in this module.

Topic Note 15.1, "Remote Sensing for Sustainable Agriculture," focuses on the different types of geographical solutions that producers and others in agricultural value chains can use to increase efficiency, reduce waste, and ultimately bring about more sustainable agricultural practices.

BOX 15.5. Key Terms Used in This Module

Remote sensing technologies are used to collect both spatial and temporal data. For comprehensive definitions of the most common types of remote sensing technologies, see Module 5.

Crowdsourcing is the process of obtaining data from a large group of people over a digital connection—for example, by broadcasting a call on the radio for farmers to send an SMS to report whether they have experienced any crop failure that season.

Crowdmapping is a subset of crowdsourcing. Data collected from a crowd are plotted onto a map using georeferencing, meaning that the data are associated with a particular point on a map. For example, farmers participating in the crowd-sourcing example mentioned above could be asked to include their location in their response so their location can be associated with their response on a map. If farmers respond by SMS, this georeferencing may have to be done manually. If respondents use a mobile app or website combined with the GPS receiver on their phones, their geographical information can be collected automatically and tagged to the information they supply with a fairly high degree of accuracy (often within a few meters). GPS receivers are available primarily on smartphones and newer feature phones; georeferencing is not possible on basic and older feature phones.

Big data analytics has an evolving definition but generally refers to mining and analyzing data for improved decision making using software and hardware (using complex algorithms and artificial intelligence, for instance) that are much more sophisticated than those used by traditional databases. Some of the techniques used to process big data are defined in Module 5.

Source: Authors.

Topic Note 15.2, "Crowdsourcing and Crowdmapping: The Power of Volunteers," focuses on how advances in social networking and data collection are enabling individuals to share hyperlocalized data in ways that have the potential to benefit society more broadly.

Topic Note 15.3, "Big Data for Analytics," focuses on how all of these agriculture-related data collected globally can be mined and analyzed in ways that lead to meaningful insights about how agriculture can be made more sustainable and productive.

Topic Note 15.1: REMOTE SENSING FOR SUSTAINABLE AGRICULTURE

TRENDS AND ISSUES

Remote sensing covers a range of technologies, many of which are described in detail in Module 5. This topic note focuses on the types of data collected by remote sensing devices and on how such data can be used to assist producers, policy makers, and researchers. In simplest terms, remote sensing refers to the use of devices to remotely monitor information from fields, grazing areas, storage containers, irrigation plots and alike, and in some cases to remotely take specific actions. For example, remote sensors are integral to precision agriculture, which aims to maximize farming efficiency and minimize waste through data to guide hyper-localized agricultural practices.

Since most remote sensors are digital devices, they collect a lot of data, and the potential impact of their data can be enhanced significantly through integration with big data analytics (see Topic Note 16.3). While remote sensors can (for instance) be used to monitor crop growth and identify anomalies, the technologies are even more powerful when they are paired with systems that can identify issues automatically and offer advice on actions to mitigate them.

Generally speaking, remote sensing devices can be classified into three types: ground, air, and space.

Ground sensors capture data based on circumstances on the ground. They can be embedded in farm equipment, such as sensors that track yield data, or they can be stand-alone devices, such as soil and water monitors, normalized difference vegetation index (NDVI) sensors, and weather stations. Some of these sensors are controlled by producers themselves, although others (portable soil sensors, terrestrial laser scanners, or weather stations) are managed by a third party. As mentioned in the overview, farmers also increasingly use smartphones as on-the-ground sensors. The sensors and GPS receivers embedded in smartphones are making it easier to collect certain types of climatic and location data.

Scientists, including breeders, are also benefiting from a new wave of ground sensors. High-throughput plant phenotyping, for example, uses a combination of spectral imaging cameras and other sensors to provide data for developing improved crop varieties (Fahlgren, Gehan and Baxter 2015; Thomasson 2015).

Aerial sensors consisted until recently of small aircraft armed with tools such as GPS, light detection and ranging (LiDAR) laser sensors, and digital still, multispectral, and thermal-imaging cameras. Flying low over a field, aerial sensors could obtain high-resolution images capable of providing farmers with information about weed growth, water stress, and even the locations of anthills (USDA 2005). This type of sensing is out of reach of the majority of the world's farmers, although it has applications for governments and larger agribusinesses. The commercial availability of unmanned aerial vehicles (UAVs, or drones) is removing some of the limitations on using aerial sensors and may eventually give many more producers, including those in middle- and low-income countries, a cheaper alternative to aircraft (box 15.6).

Even more removed from the field are **space-based satellites**. Any ground or aerial sensor that uses GPS relies on satellites to calculate its positioning. Satellites also have a range of other imagers and sensors. Different spectral bands can capture different types of information. For instance, thermal infrared can measure surface temperatures, whereas green can be used to assess plant vigor. Spectral bands also provide different levels of information about specific physical objects that are seen. Image 15.1 illustrates how different spectral bands provide different levels of granularity about trees, ranging from identifying an object as a tree with panchromatic (black-and-white) imagery to indicating what class of tree it is and even identifying the specific species with short wave infrared (SWIR) imagery.

Not all satellites capture data across all spectral bands, and only some satellite data are made publicly available (for instance, Landsat data). A satellite or a constellation of satellites will also vary in the frequency with which complete coverage of the globe can be provided, which in turn affects how frequently they capture new data on a specific location. For example, Landsat 8 and Landsat 7 take 16 days to cover the globe. Microsatellite firms, such as Planet Labs, on the other hand, aim to provide daily coverage of the world's entire landmass.

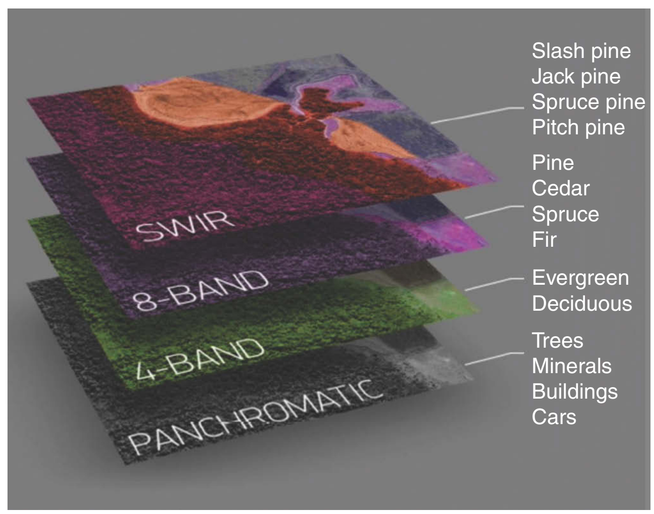
In many cases, certain levels of skill and expertise are still required to analyze and make sense of satellite data. This situation is changing, however. Advances in satellite technologies and increased competition are making it possible for

BOX 15.6. The Future of Drones for Smallholders

Will drones enable remote sensing to become a less remote prospect for smallholders? The International Potato Center received a grant from the Bill & Melinda Gates Foundation in late 2014 for local institutions to pilot the use of drones for remote monitoring of sweet potato fields planted by smallholders in Tanzania and Uganda.[a] In addition, new firms such as Raptor Maps in the United States and CI-Agriculture in Indonesia are combining big data analytics with data from drones and other remote sensors to provide farmers with precise information on their farms. It is too soon to tell whether either firm will succeed, but they are a bellwether of what is to come.[b]

Source: Authors.
a. Quinn 2014.
b. Lavars 2015; Freischlad 2015.

IMAGE 15.1. Different Spectral Bands for Satellite Imagery



Source: DigitalGlobe.

even small-scale producers to receive insights from satellite sensors that are pertinent to their farms, or at least to their general geographic area, directly on their mobile phones. Farmers generally do not access the satellite data directly; they subscribe to a mobile service that makes use of it.

For example, coffee farmers in Rwanda can use the WeatherSafe app, which combines satellite and farm-specific data, to receive localized weather information and farming recommendations (European Space Agency 2014). While eLEAF, which is based in the Netherlands, uses a combination of meteorological and remote sensing data to monitor crop growth and water use on farms in real time (see the innovative practice summary below). Another interesting example of expanding access to satellite information is the partnership between the Food and Agriculture Organization (FAO) and Google to increase the accessibility of its geospatial tracking and mapping products (Graziano da Silva 2016). One recent output from this partnership is the Water Productivity Open-access Portal (WaPOR), which currently has 250m spatial resolution data on agricultural water productivity for all of Africa and the Near East.[12]

The trend toward increasingly sophisticated and in some cases miniaturized remote sensors will undoubtedly continue in the coming decades, and prices for many of them will almost certainly continue to fall. It is estimated that the precision farming market will see a compound annual growth rate of over 13 percent from 2015 through 2022 (BIS Research 2014).

It is important to understand that even though remote sensing technologies are becoming more accessible, access to those technologies will not be distributed evenly. The majority of producers, especially smallholders, are likely to benefit from remote sensors that are built into or are accessible via mobile phones, although their use will depend partly on whether producers have access to smartphones capable of using the technology. GSMA Intelligence calculated that 43 percent of people in developing countries owned a mobile phone in 2014, a figure that they expect to grow to only 55 percent by 2020. Over that same period, however, GSMA anticipates that smartphone ownership will grow from around 25 percent to over 60 percent.[13] Those trends are both positive but clearly show that significant portions of the population will remain without direct mobile access.

Possibly other types of basic ground sensors and potentially even small drones could find their way into the hands of poor, small-scale producers. In the short term, however, it remains likely that most of the world's farmers will have access to these types of devices at best through an intermediary.

Researchers and policy makers potentially have much to gain from advances in remote sensing, which should significantly increase the amount of precise agricultural data available to them (box 15.7). These data, combined with big data analytics (discussed in Topic Note 15.3), could drive the future of agricultural research and also guide the design of more precise interventions by governments and development organizations seeking to support specific subsets of producers (poor smallholders, transhumant livestock producers, women diversifying into horticultural crop production, and so on). The potential for increased precision at the farm level to improve the efficiency of resource use will be particularly important for the ability of all producers and for all types of agriculture to adapt to climate change.

Of course, the policy implications of some of these technologies must be addressed first to ensure that benefits can be gained from them without undermining broader public interests. For instance, many countries have yet to resolve legal issues related to the use of personal drones for agriculture. Some have opted to control personal drone use from a public nuisance perspective, while others approach the issue from a national security perspective. Many governments also restrict the use of satellite imagery. Until recently, the U.S. government restricted the sale of satellite images at resolutions sharper than 50 centimeters. Those restrictions have been eased somewhat, and now black-and-white images up to 25 centimeters in resolution and color images up to

BOX 15.7. Advances in Remote Sensing

At the ICT for Ag conference in Washington, DC, in 2015, presenters noted that a couple of trends in remote sensing for agriculture could greatly unlock the value of their data (Zoltner 2015):

- Free and open source platforms for interpreting the data are becoming more available.
- High numbers of low-cost sensors can match the accuracy of high-end sensors through a greater number of data points.

Source: Authors.

12 WaPOR can be accessed online at http://www.fao.org/in-action/remote-sensing-for-water-productivity/database/database-dissemination-wapor/en/.

13 Accessed on https://gsmaintelligence.com/. All data © GSMA Intelligence 2015.

1 meter in resolution may be sold to non-U.S. customers (Ferster 2014).

Measuring the impact of remote sensing for sustainable agriculture depends entirely on the type of sensing that is used and its purpose. Broadly speaking, though, farmers using remote sensors are anticipating seeing improvements in productivity and/or reductions in costs, not to mention a reduction in negative effects on the environment.

LESSONS LEARNED

Until recently, even in high-income economies, full use of the spectrum of remote sensing technology required tens of thousands of dollars in investment, meaning that large numbers of farmers often could not benefit. In low-income countries, producers' use of remote sensing remains in an extremely nascent phase. The world is embarking on a new era of more affordable remote sensing, however. In moving forward, policy makers, researchers, and practitioners can benefit from applying some lessons gained from using these technologies to date.

Clarify Legal Limbos

Precision agriculture technologies, including remote sensing, will remain on the cutting edge of technological capabilities for some time. As new technologies emerge or previously restricted technologies are made commercially available, governments will need to move quickly to address any legal issues that could impede their use. The legal uncertainty surrounding drones is the current case in point, although undoubtedly other technologies will find themselves in the same situation.

Expand Digital Literacy

Remote sensing relies on a variety of digital hardware and software. Producers, especially poor producers, will benefit from these tools only if they understand how to use them. Fostering digital literacy is potentially no small feat, and may require continuous training and support for some groups of producers. In the short term, since many producers will not have direct access to remote sensing tools, there is an opportunity to acquaint them with these tools via an intermediary, such as an extension agent or lead farmer. One such example of strengthening intermediary capacity comes from the partnership between FAO and Google mentioned earlier. As part of that partnership, 1,200 FAO staff and partners will receive trusted tester credentials on the Google Earth Engine, as well as receiving training on how to use it. In turn, FAO will train its staff and technical experts in member countries on how to use this technology (Graziano da Silva 2016).

Expand the Capacity to Interpret Data

The next step beyond understanding how to use remote sensing tools is to understand how to interpret the resulting data and translate those insights into action. The data tend to be highly technical and potentially difficult for someone with a basic level of education to understand. Any remote sensing tool that aims to benefit such producers will need to closely consider this reality during its design.

Ensure Interoperability and the Adoption of Standards

A major impetus for creating the Open Ag Data Alliance was the lack of interoperability between precision agriculture technologies and data standards. The Open Ag Data Alliance was particularly concerned with the fragmentation of data across the different devices and platforms used by a typical U.S. farmer.[14] The lack of application program interfaces (APIs) connecting different software applications prevented producers from easily seeing all of their data in one place. As development organizations begin to promote the use of remote sensing technology among producers, especially disadvantaged producers, it is crucial to avoid these pitfalls by partnering with providers that adhere to open standards and interoperability—or to encourage providers that do not adhere to open standards to do so.

Strike an Appropriate Balance between Privacy and the Public Interest

This lesson is explored in more detail in the Lessons Learned section of Topic Note 16.3 on big data. In short, while access to all of the hyperlocalized data generated by ground and aerial sensors could be of huge value to researchers, it may not be in producers' best interest to share those data. While providers of remote sensor software may embed their right to sell user data to third parties in their terms and conditions, practitioners should ensure that producers are asked for explicit informed consent to do so. In some cases, this approach will mean resisting the temptation to enable individuals to opt in by default in order to ensure a rich data set. At the same time, researchers, practitioners, and policy makers should promote strong data privacy standards to

14 Two figures on the Open Ag Data Alliance website clearly illustrate the current disconnectedness of remote sensing systems and the envisioned future interoperable state; see http://openag.io/about-us/principals-use-cases/.

protect the personal data of producers, including their farm- or enterprise-specific data. In some instances, it may be necessary to explore standards that promote unlinking farm-related data from exact geographic coordinates—providing them at the village or district level instead—to balance producers' interests in protecting their specific farm data with researchers' and policy makers' desires for access to some level of aggregate information to use in their research and decision making.

INNOVATIVE PRACTICE SUMMARY
Using Multispectral Satellite Images and Energy Surface Balance Models to Calculate Crop and Water Productivity

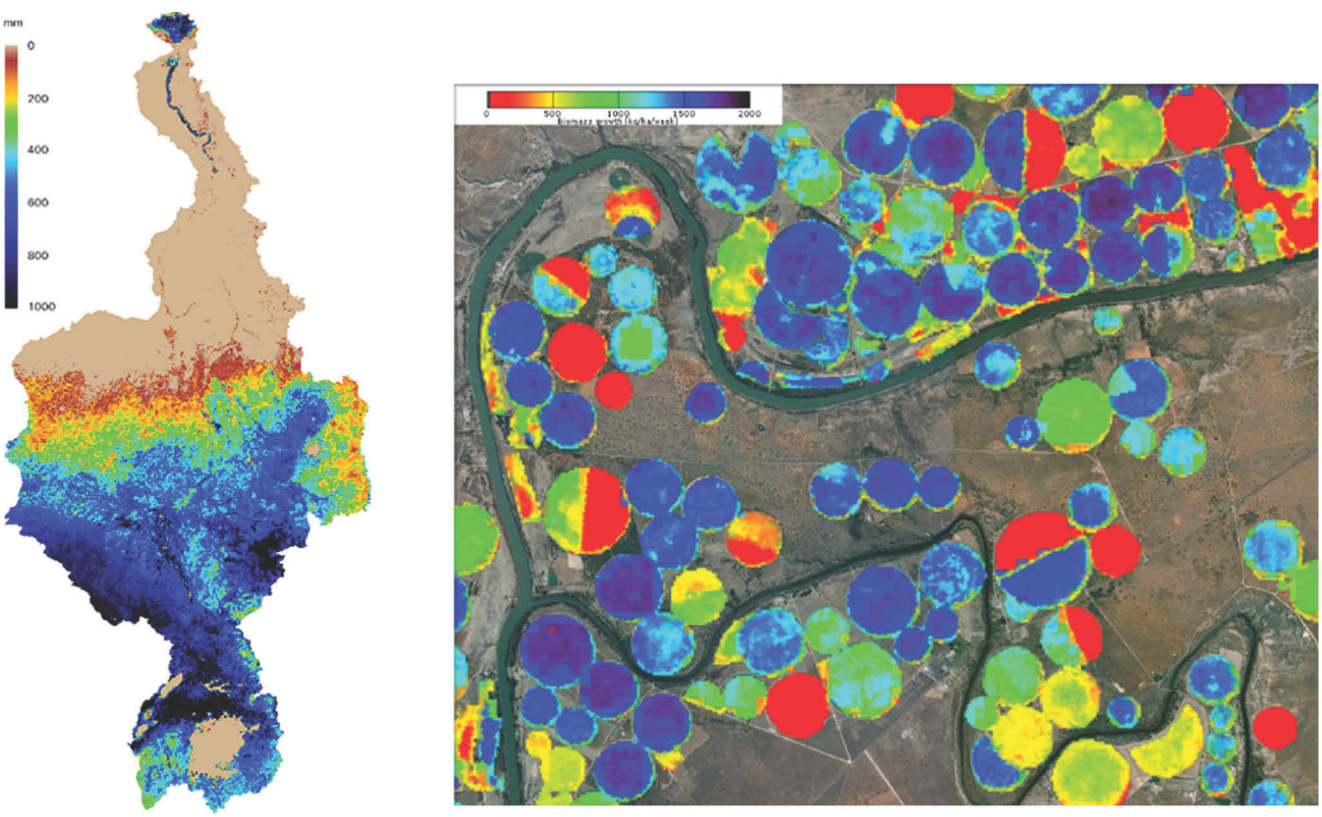
The Netherlands-based firm eLEAF has developed a series of algorithms that use a combination of meteorological and remote sensing data to monitor crop growth and water use on farms in real time. A technology called Pixel Intelligence Mapping (PiMapping) delivers data at a range of spatial resolutions (image 15.2). Widely used are 250-meter data and 20–30-meter data; however, with the launch of the European Union Sentinel satellites in 2015, resolutions as high as 10 meters are within reach. The lower-resolution data can be used for monitoring large areas at a regional or national scale, such as entire river basins. The higher-resolution data allow identification of individual fields to monitor crop growth changes on individual farms.

Using eLEAFs FieldLook-platform, individuals can access a range of PiMapping data, including biomass production, evapotranspiration, transpiration, evaporation, biomass water productivity, leaf area index (LAI), NDVI, and fractional vegetation cover (FVC). In addition, by delivering both historical and real-time data, users can use PiMapping data to monitor changes over time as well as to check on current conditions.[15]

In Sudan, the Hydraulics Research Center (HRC), with funding from CTA, used eLEAF's technology to monitor the Gezira Irrigation Scheme, which is one of the world's largest

15 Information taken from eLEAF's Company Profile on its website (http://www.eleaf.com/) and confirmed by eLeaf staff.

IMAGE 15.2. Examples of Low- and High-Resolution Pimapping Data



Source: eLEAF.
Note: Images are used with permission; further permission required for reuse. On the left: Evapotranspiration in mm/year for the entire Nile Basin at a resolution of 250m; on the right: 30m resolution biomass production makes growth variations within individual grain fields visible in South Africa.

irrigation schemes. Specialists from HRC used the data to send registered farmers SMS notifications to let them know the best time to irrigate their fields and to apply fertilizer. The service was piloted with 44 farmers during the 2014/15 planting season. All of the farmers increased their productivity as a result of participating in the pilot. One farmer, for instance, saw wheat production increase to 12 sacks per acre, up from 3 sacks per acre in the previous year (CTA 2015).

Helpful resources

- Remote Sensing Technology Trends and Agriculture, https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/31/DG-RemoteSensing-WP.pdf.
- International Society of Precision Agriculture, https://www.ispag.org/.

Topic Note 15.2: CROWDSOURCING AND CROWDMAPPING: THE POWER OF VOLUNTEERS

TRENDS AND ISSUES

This topic note explores how volunteers are contributing agricultural data to larger and broader data sets, as well as helping to ground truth the accuracy of maps and underlying data sets. It explores the implications of this practice for individual producers as well as for implementers, researchers, and policy makers. In particular, it provides a perspective on how organizations are capitalizing on ICT-enabled approaches to collect data from a wider population than would have been feasible using traditional data collection methods, such as surveys conducted by enumerators. The note also examines how organizations are using those same ICT applications to share insights gleaned from analysis of the raw data with the respondents who provided that data.

The term "crowdsourcing" was coined in *Wired* magazine in 2006 but its origins may date as far back as the early 18th century, when the British government put out an open call to find a reliable method to calculate the longitude of a ship (Dawson and Bynghall 2012). Crowdsourcing in the modern sense of the word tends to refer to information that is collected or tasks that are performed by a large group of people, often in different locations, via some form of digital device (such as a computer or mobile phone). Crowdsourced data are collected primarily via SMS, websites, mobile apps, social media, email, and voice calls. In the context of development, crowdsourcing's rise to global prominence can perhaps be dated to early 2008, when a group of technologists and activists developed a platform, dubbed Ushahidi, to track outbreaks of violence in postelection Kenya. Ushahidi was not the first use of crowdsourcing in development, but the international attention it garnered helped many more people to realize the potential of this approach.

From a data collection perspective, crowdsourcing can be combined with georeferenced data (often referred to as crowdmapping), which often come from GPS coordinates collected via a mobile phone—although, in some cases, the location is reported by the respondent. Through this pairing, an individual with access to the data can see exactly where in the world these data came from. Another advantage is that because these data can be processed in real time by data collection platforms, the parties collecting the data can access them almost immediately. The only delay occurs when data are input into a mobile data collection tool in offline mode, when the phone is out of network range. In those instances, the data will not be synchronized to the main database until the phone is connected to the Internet again.

Crowdsourcing can also be used in combination with geospatial images, calling on the power of the crowd to help identify changes to a specific location. For example, DigitalGlobe released images before and after Typhoon Haiyan struck the Philippines in 2013. The images covered 100,000 kilometers. More than 4,600 individuals tagged over 400,000 items, including 143,155 damaged buildings and residences.[16] This example is not specifically related to agriculture, but it clearly demonstrates the power of crowdsourcing and crowdmapping.

It is easy to see why this approach would appeal to anyone who needs data from a wide geographic range. Unlike traditional data collection methods that require enumerators,

16 To learn more, see https://www.digitalglobe.com/sites/default/files/Crowdsourcing-DS-CROWD.pdf.

often using paper forms, to be on the ground, crowdsourcing enables the collection of data from a significantly larger number of individuals, across a broader geographic area, in much less time and for less money. Crowdsourcing services can also be designed to be bidirectional, giving individuals, even if they have not contributed data, near-instantaneous access to the aggregate data and analysis. This immediacy is a positive shift from traditional methods of data collection, whereby it took months for data collected from communities to make it back to those communities, if at all.

Crowdsourcing is predicated on the fact that the target population has some form of digital connection—most likely via a mobile phone. As discussed in the overview, the reality is that only slightly more than two-thirds of the world's adult population owns a mobile phone.[17] Nor does simply owning a phone mean that an individual will use it to provide information to a crowdsourcing initiative. Differences in literacy, numeracy, and digital literacy all affect the ability to use a mobile phone effectively. Users also need access to sufficient electricity[18] to keep a phone charged and sufficient income to purchase airtime to keep the phone number active.

Given these variables, the feasibility of crowdsourcing will depend on the local context. It is also important to be mindful of those without access to ensure that their opinions or inputs are included—or that if they are excluded, those limitations of the data are clearly stated.

Assuming that the local context is conducive to crowdsourcing, the approach can be used in agriculture in a number of ways. Some of the most likely uses are as follows:

- **Tracking pest and disease outbreaks**. Delays in traditional pest and disease reporting often prevent the authorities from taking decisive action to contain outbreaks. By crowdsourcing information on the incidence of pests and diseases, governments and researchers may identify outbreaks before they spread and take action accordingly. In 2012, Zambia's Disaster Management and Mitigation Unit created a crowdmap to track armyworm sightings, which could be submitted by SMS or voice calls using a short code. Given this information, the government was able to target resources and contain the outbreak (Silversmith and Tulchin 2013).[19] A similar approach was used in Uganda to monitor banana bacterial wilt using Ureport, an SMS-based polling service (Bujoreanu 2013).
- **Collecting local weather information.** For weather forecasts to be useful to producers, they need to be hyperlocalized to their farms. Traditional weather forecasts, based on a blend of satellite data and ground-based weather stations, generally cannot provide that level of specificity. The next generation of weather forecasts is taking advantage of sensors built into smartphones to collect extremely localized weather information. Apps such as PressureNet, Sunshine, and WeatherSignal all provide such services, although primarily for urban users. Eventually, these types of services will be practical for most smallholders.
- **Collecting market prices.** Market information services (MIS) have long relied on enumerators reporting daily prices for select markets. Services like AGROAM are experimenting with the possibility of indirectly crowdsourcing market prices by offering a buyer-seller matching service. By taking an average of what prices are paid for specific crops in specific areas, they are able to provide a snapshot of actual market prices. This method is highly dependent on volume, since low volumes of transactions for certain crops or in certain areas could skew the actual market price average. Esoko is doing something similar with its MarketPlace app, taking price data from purchases made through the system to show price trends over time.
- **Facilitating access to markets.** Among a number of other mobile services, the Connected Farmer Alliance, highlighted in the innovative practice summary below, is using mobile phones to enable smallholder farmers to market their produce to prospective buyers in Kenya, Tanzania, and Mozambique.
- **Agriculture knowledge sharing.** Agricultural knowledge has actually been crowdsourced for decades,

17 This figure was roughly calculated at 69 percent as of June 1, 2015, based on total unique mobile subscribers from GSMA Intelligence divided by the total global population over the age of 15, which was calculated from the Population Reference Bureau's 2012 data and U.S. Census Bureau data. As young people under the age of 15 do own mobile phones, this estimate is likely to be higher than the real number of adults with mobile phones.

18 Some estimates state that by 2030, almost 900 million people globally will still lack access to electricity (Practical Action 2012).

19 The crowdmap from that outbreak can be found at https://armyworms.crowdmap.com/main?l=en_US.

in the form of radio call-in programs on which farmers direct their questions to experts. With mobile phones and social media, new channels are available for individual producers to pose questions and receive responses from experts as well as their peers. One example is the Awaaz.De interactive voice response platform, which has been used by organizations in India to provide agricultural information to farmers. The service allows callers to record questions for experts to respond to, listen to questions and answers from others, and also to record their own responses to questions.

- **Facilitating land administration.** Indigenous communities in some parts of the world have fallen victim to land grabs by outsiders because they have not been able to demarcate their traditional rights to land on a map. The Rainforest Foundation UK has been helping communities in the Congo River Basin use GPS-enabled mobile phones to map the land that they use for hunting and gathering. Module 14 contains more information on crowdsourcing for land administration.
- **Crop and livestock monitoring.** One method for crowdsourcing crop and livestock data is to ask producers to submit information about the crops that they are growing and their livestock numbers. Results of such efforts have been uneven. For instance, the Mauritius Breadfruit Sector Consortium tried to map all of the island's breadfruit trees with less than impressive results—fewer than 70 out of an estimated 3,000 trees were mapped (Hosenally 2012). A more globally focused example, highlighted in the innovative practice summary below, is Cropland Capture, a game developed by Geo-Wiki that shows players a satellite image and asks them whether they see any cropland (Gustafson 2013).
- **Conducting research.** As noted, the proliferation of mobile phones offer researchers opportunities to survey farmers without sending enumerators to meet them in person. Short polls can be conducted via SMS, while voice, mobile apps, and websites can be used for longer polls, particularly those requiring qualitative responses. GeoPoll's Food Security Service is one example of a service that crowdsources data for research. The service has a database of 200 million users in roughly a dozen countries and currently offers surveys to capture data on two indicators: Food Consumption Score (FCS) and Reduced Coping Strategies Index (rCSI).
- **Monitoring food security.** The UN Global Pulse has used numbers of tweets to track inflation in the rice price (Crimson Hexagon 2011). Although this type of social media sentiment analysis may not always be effective for determining whether changes in food security are occurring, it is an area worth further exploration.

How the success of a crowdsourcing initiative is measured will vary depending on how and why crowdsourcing is being used. At the output level, the number of people engaged in the crowdsourcing effort and the total number of contributions made can be tracked. At the outcome level, two key indicators can help to measure the success of a crowdsourcing initiative. The first is the impact of crowdsourced contributions on the overall outcome being tracked. For instance, if the overall project aims to increase farmers' knowledge of a particular topic, what, if any, changes to that indicator can be attributed to farmers' access to crowdsourced resources? Another method of measuring outcomes is to compare the cost versus benefit of crowdsourcing and traditional methods. When using crowdsourcing to conduct research, ask how the cost of crowdsourcing and the benefit derived from it compare to the costs and benefits of traditional methods of conducting research.

LESSONS LEARNED

It is easy to get excited about crowdsourcing. This approach potentially facilitates engagement with a large and dispersed audience, often at a fraction of the cost of traditional methods of engagement, and can deliver information in near-real time.

As great as all of this sounds, experience from the field reveals some challenges in using crowdsourcing effectively. Thankfully, good practices are emerging from the experiences of practitioners, including lessons on specific challenges and specific actions that can be taken to mitigate them.

Data Quality

Since crowdsourcing tends to entail the collection of information from people who are not experts on the subject, the likelihood of error is perhaps higher than what one might expect from data that are input by trained enumerators. Sometimes participants can be intentionally misleading; for example, AGROAM, which shares market price information, encountered individuals who were deliberately trying to game the system for their own benefit (Conor 2014).

One approach to reducing data input errors is to design and test the platform rigorously with the target audience to ensure that the platform is intuitive and easy for them to use. Methods such as user-centered design can be helpful for doing this. IDEO.org's *Field Guide to Human-Centered Design* is a great resource for learning how to employ this approach.

It is a bit more complicated to protect against people who are gaming the system or simply sharing incorrect information that they believe to be true. This issue is truly a challenge only when the number of contributors is small. With a large enough pool of contributors, such outliers can be identified and discarded—assuming that there are no large group biases. For that reason, the best protection against gaming is to have a large pool of contributors who can validate and crosscheck each other. Some technology providers have also developed algorithms to determine the reliability of contributors in order to weight inputs, such as DigitalGlobe's CrowdRank.

During the design phase, it is advisable to think about all of the reasons why someone would submit false or incorrect data so that checks can be devised for them. Devising a clear data verification process will also help to mitigate data quality issues.

Accessibility

Because not everyone has access to mobile phones or the Internet, a subset of the population will probably be excluded from any crowdsourcing initiative. The issue of limited accessibility is not one that any single project will be able to fix. At a minimum, however, it should be possible to learn about and understand the access of the target audience. Remember to ask such questions as: What groups of people have less access than others? How dependable is people's access (to electricity, network coverage, airtime)? How can access be extended to those without it (e.g., can the project promote device sharing)?

User Capacity

Participants need to know how to use the crowdsourcing platform or service properly to input data. In the Mauritius breadfruit mapping initiative mentioned earlier, one of the main reasons for people's failure to contribute was their limited ICT skills. Limited skills also reduced data quality, as users placed some trees in the sea or in the middle of the street (Hosenally 2012).

Understanding users in advance will help to mitigate this challenge to some extent. Designing a platform or service aligned to users' current capacity, as opposed to one that depends on massive amounts of user training, is always preferable. It will rarely be possible or cost-effective to train large amounts of people to use the platform. It also helps to design checks into the platform to catch errors and notify users of the correct way to do something.

User Interest/Incentive

Crowdsourcing initiatives can succeed only if they attract a sufficient number of people to participate. Limited participation will likely yield data of limited value for analysis. If users see no value in contributing, they are less likely to do so. In some cases, users may have an inherent interest to contribute. In other cases, some level of incentivization may be necessary. The incentive can be something as basic as offering points and badges in some cases, although it may also extend to small financial rewards, particularly if contributors need to use their own airtime to participate. Younger, more tech-savvy producers may also be more likely to participate than older producers.

Again, start by developing an understanding of the target users. What are their interests? What are their needs? What might incentivize them to participate? The *Field Guide to Human-Centered Design* has good activities for learning about users, both in the mainstream and at the extremes, to design incentive systems and marketing campaigns appropriately.

Protecting Privacy

Since individuals will be asked to share information, it is important to consider how their privacy will be protected. Privacy becomes an even more important issue when crowdsourced data are georeferenced and can reveal where people live. When using a third-party provider, be clear on how they use the data as well.

The U.S. Federal Trade Commission report, "Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policy Makers," is a useful resource for starting to consider these challenges. Questions to ask, particularly when working with third-party providers, include: How will the provider secure disaggregated data? Will the provider sell data to other parties? If so, how will the provider notify users that their data are being sold, and will users be compensated in some way?

Unintended Consequences

One of the benefits of crowdsourcing is its potential to democratize data and increase transparency, yet it is also possible that the data can be used for nefarious purposes. For example, an unscrupulous government or company armed with detailed information about a marginalized community might be able to use that information to take advantage of them. It is not farfetched to imagine a scenario in which smallholder farmers who have reported higher-than-normal productivity via a georeferenced crowdsourcing campaign find themselves in the sights of land grabbers (McLaren 2015).

It is difficult to think about what types of unintended consequences might arise. That said, it is important to try to identify potential risks, such as the misuse of any data that are made available, and to develop mitigation strategies based on the likelihood of those risks occurring.

Helpful resources

- Ushahidi Toolkits include a wealth of advice and guidance, including 10 questions to ask before starting a crowdsourcing initiative see https://wiki.ushahidi.com/display/WIKI/Ushahidi+Toolkits
- Crowdsourcing Applications for Agricultural Development in Africa, http://pdf.usaid.gov/pdf_docs/PA00J7P7.pdf
- ICT Update: Crowdsourcing and engagement, http://wayback.archive-it.org/3908/20150511072324/http://ictupdate.cta.int/(issue)/69

INNOVATIVE PRACTICE SUMMARY
Crowdsourcing Supplier Data via Mobile Phone

The Connected Farmer Alliance is a public-private partnership between Vodafone, the U.S. Agency for International Development (USAID), and TechnoServe focused on promoting commercially sustainable mobile agriculture services for smallholder farmers in Kenya, Tanzania, and Mozambique.[20] The Alliance pilots initiatives aiming to create a better ecosystem for mobile services in the agricultural sector, affecting production throughout the supply chain.

Data Generation

One of the main focus areas of the Connected Farmer Alliance involves enterprise solutions that enable enterprises to better source from small farmers and allow farmers better access to markets. The data are gathered and distributed through a suite of modules, including a registration module allowing an agent of an enterprise to register a farmer who supplies a particular type of produce. Farmers may also register themselves as suppliers. In this way, the service enables the remote gathering of crowdsourced data to identify who and where farmers are and what types of crops they produce. The data are highly structured. They are referenced temporally and spatially and clearly identify individuals so that participating enterprises can distinguish specific farmers and their products. The typical participating enterprise is a mid-sized national company that sources produce from small farmers and seeks more detailed data and interaction with available suppliers.

Building upon the crowdsourced supplier data are a series of additional modules, including two-way communication that enables enterprises to share information with, or survey, farmers. A receipting module, integrated with M-PESA (a mobile money service), allows enterprises to send receipts and pay farmers at the point of sale, identifying the time, price, and volume of the purchase, which increases transparency. Another module allows enterprises to offer short-term loans through M-PESA, enabling cash advances that are later deducted from payments for produce. The enterprise can use the data generated through the registration and receipting modules to assess farmers' creditworthiness, something that was not previously possible for the majority of agribusinesses. A tracking module enables enterprises to better track collection processes and points to streamlined product collection. At this time, the size of the crowdsourced data set does not yet approach big data, but Vodafone is bringing this first suite of modules to commercial markets for much broader deployment.

Data Interpretation

Vodafone works with its subsidiary, Mezzanine, on the development and management of the data collection platform, which is locally hosted in the Kenyan, Tanzanian, and

20 This innovative practice summary originally appeared in the World Bank's report *Big Data in Action for Development*. The original case study can be found on pages 39–41 of that report, which is available at http://live.worldbank.org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version.pdf. It was slightly updated for this module based on inputs from Drew Johnson, interim regional program director of the Connected Farmer Alliance, which were provided in February 2016.

Mozambican markets and protected by high-level security mechanisms. Data are available only to the enterprise and participating farmers. For the surveys, enterprises receive only aggregated responses, not individual records. Vodafone is working with enterprise customers on the most convenient way for farmers to submit data while ensuring confidentiality for them and for businesses. The details of data privacy will be governed by Vodafone's data privacy policies to ensure ongoing protection.

Within the Connected Farmer Alliance partnership, TechnoServe is charged with analysis and interpretation of how the modules are performing for the enterprises and farmers. Insights are currently being gathered through traditional survey methods. Those methods include assessing goals for the participants at the outset of the project, determining areas of measurement, and collecting input through questionnaires during the process. Additionally, the Connected Farmer Alliance supports enterprise partners in their own data analyses of information and outcomes.

Insights for Action

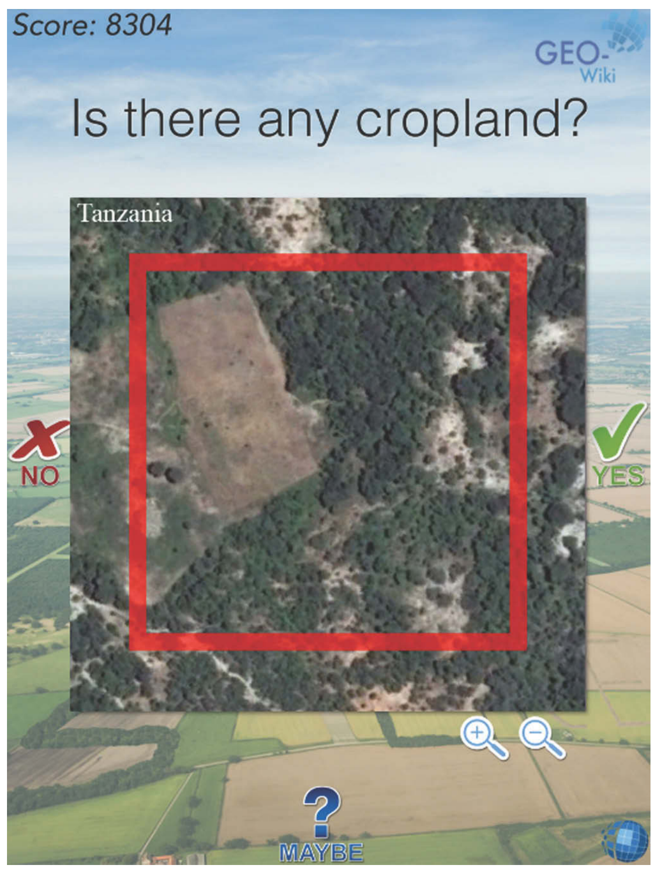
Although enterprises are just starting to adopt the technology, some insights are emerging into the benefits of the modules. Farmers who receive M-PESA for loans and payments reduced their costs by avoiding expensive, time-consuming, and risky trips to the enterprise office to collect cash. The receipting module has reduced the costs of enterprises by increasing their operational efficiency and transparency. A key benefit of mobile solutions for farmers is the increased access to information. It is difficult to make generic content services meaningful to small farmers whose local realities may vary significantly within a distance of just a few kilometers. The targeted information flow permitted by the two-way information module appears to provide information that is particularly relevant to the stakeholder farmers and to enhance the face-to-face interactions among farmers and enterprises.

INNOVATIVE PRACTICE SUMMARY
Combining Gaming and Crowdsourcing to Identify and Monitor Cropland

The Geo-Wiki project was established by the International Institute for Applied Systems Analysis (IIASA), the University of Applied Sciences Wiener Neustadt, and the University of Freiburg (Germany) in 2009 to try to address one of the main challenges with current global land cover data sets—the large discrepancies between them. The project asks a network of volunteers to help identify, among other things, human impact in satellite imagery of land cover (see image 15.3).

A good understanding of the location of the world's cropland is important for a number of reasons, including identifying where the best investments could be made to increase production (Gustafson 2013). The good news is that experts are not needed to identify cropland accurately, as research by IIASA has revealed. A comparison of 53,000 data points input by experts and nonexperts found that the nonexperts were just as good as experts at identifying human impact (specifically, cropland) in satellite images (See 2013). A comparison of the accuracy of crowdsourced cropland maps to three major cropland maps (GLC [Global Land Cover]-2000; Moderate Resolution Imaging Spectroradiometer, MODIS; and GlobCover) in Ethiopia found the crowdsourced data to be the most accurate (See et al. 2013).

IMAGE 15.3. Screenshot of Cropland Capture



Source: Geo-Wiki Project.
Note: Used with permission. New request for permission is required if image is reused.

With this information in hand, Geo-Wiki launched a mobile and Web-based game called Cropland Capture that transforms anyone in the world into a citizen scientist. The concept is quite simple. Players of the game are shown an image and asked if it shows cropland at a certain site (answering "yes," "no," or "maybe") (figure 16.4). Images contested by players are sent to an expert to make the determination. The resulting crowdsourced data have been used to create a new global cropland map that Geo-Wiki hopes will be more accurate than other available sources.

The game was launched in 2013, followed by 25 weeks of competition rewarding the top three players of the week with prizes to incentivize participation. Although the game is still available for download, the lessons learned from the game have been used to develop a more generic version of Cropland Capture called Picture Pile, which is being used to gather information on evidence of deforestation.[21]

21 More information on Cropland Capture is available at http://www.geo-wiki.org/oldgames/croplandcapture/. Picture Pile can be found at http://geo-wiki.org/games/picturepile.

Topic Note 15.3: BIG DATA FOR ANALYTICS

TRENDS AND ISSUES

It is perhaps only mildly hyperbolic to state that big data will have a dramatic impact on the 21st century economy. In 2014, the big data market generated more than US$27 billion in revenue from the sale of hardware, software, and professional services. Within the next 10 years, it may generate more than US$80 billion in global revenues (Kelly 2015). Even that figure pales in comparison to the potential revenue that might be generated and costs saved from the application of big data. The McKinsey Global Institute, for instance, estimated in 2011 that big data could have a value of more than US$1 trillion annually from just three sectors: the U.S. health care system, European public sector administration, and personal location data (McKinsey Global Institute 2011).

> Within the next five years, big data will become the norm, enabling a new horizon of personalization for both products and services. Wise leaders will soon embrace the game-changing opportunities that big data affords for their societies and organizations, and will provide the necessary sponsorship to realize this potential. Skeptics and laggards, meanwhile, look set to pay a heavy price.
>
> —"Big Data Maturity: An Action Plan for Policy Makers and Executives." Chapter 1.3 in *The Global Information Technology Report 2014* (World Economic Forum and INSEAD).

This topic note examines how agriculture could make use of big data and discusses what has been learned so far. Topic Notes 16.1 and 16.2 focus on sources of data—remote sensors and crowdsourcing—whereas this topic note focuses on how all of these data can be more effectively analyzed and acted upon.

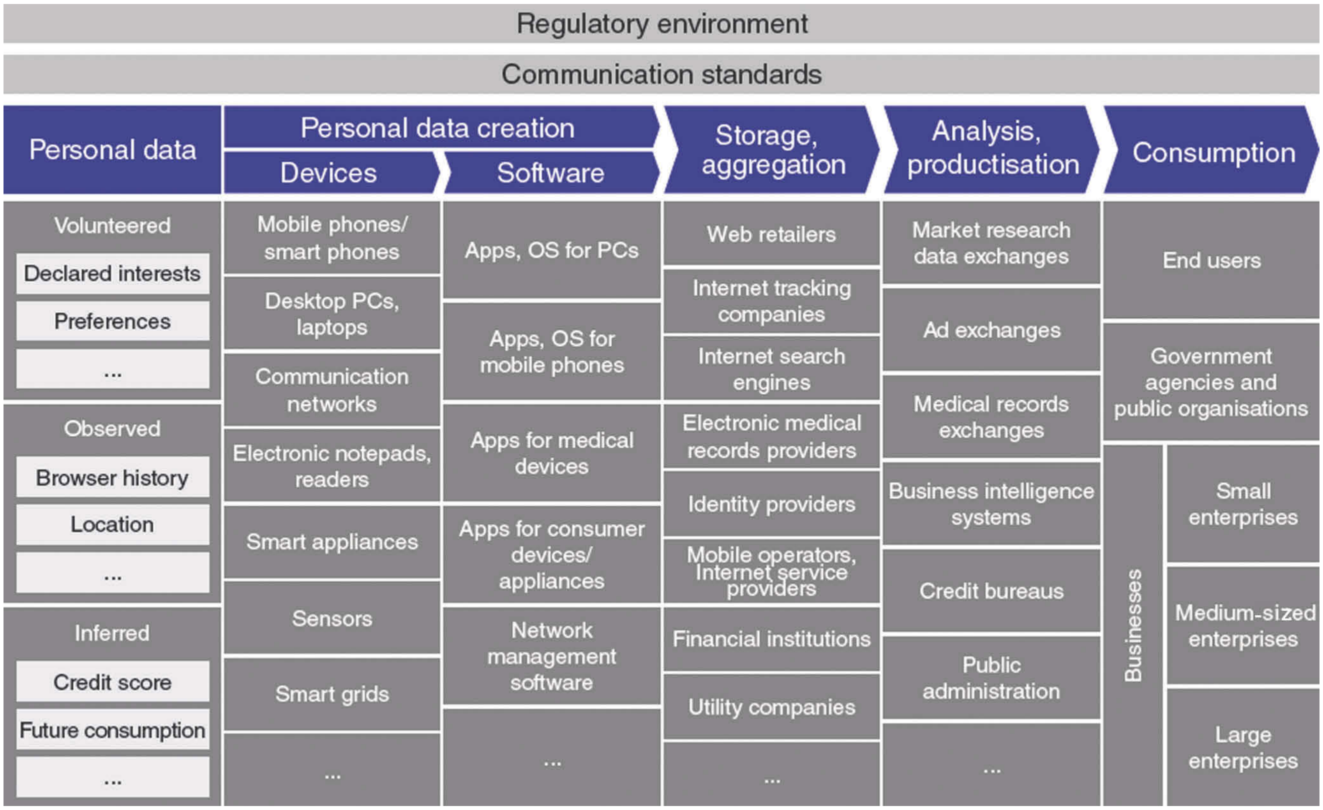
As mentioned in the overview, no widely accepted standard definition of big data exists, although many people talk about big data as having the following characteristics, referred to as the "3 Vs," defined in 2001 by analyst Doug Laney (Laney 2001):

- **Volume** refers to the sheer volume of digital data being produced globally.
- **Velocity** refers to the speed at which data can be captured and analyzed.
- **Variety** refers to all of the different types and formats of data being produced.

Since then, some people have opted for additional Vs. For instance, SAP Business Innovations has added **veracity**, which refers to the quality of the data, and **value**, which refers to the potential business value derived from it (Saporito 2014).

Figure 15.2 shows the full life cycle of data from creation to consumption. Numerous types of digital devices and software amass data from a variety of places, including individuals, the public sector, and the private sector. Data can be volunteered (explicitly shared), observed (captured by recording actions, such as Web browsing history or call detail records, CDRs), or inferred (based on an analysis of volunteered and/or observed data) (World Economic Forum 2011). The data are stored and aggregated by a range of entities, including websites, mobile network operators, and development organizations. The next step is analysis, which is sometimes done by the same entities that store and aggregate the data and sometimes by third parties. A growing number of companies offer what is known as BDaaS (Big Data as a Service). These providers offer access to analytics tools and services for a fee, as opposed to firms that sell insights based on analysis they conduct.

FIGURE 15.2. Life Cycle of Digital Data



Source: Bain & Company.[22]
Note: Used with permission. New request for permission is required if image is reused.

The last step in the chain is the consumption of the insights and intelligence gleaned from big data analytics; the consumers can include governments, businesses, researchers, and sometimes even individuals.

Broadly speaking, big data can be used to support six areas with potential applications in agriculture:

- **Awareness—**learning that something is happening by using sentiment analysis to identify potential trends in people's opinions or concerns. For example, as mentioned, the UN Global Pulse roughly tracked the inflation of the price of rice in Indonesia by analyzing tweets on Twitter.
- **Understanding—**learning why something is happening, such as why food prices have risen or why water shortages have arisen.
- **Advice—**providing targeted and specific advice, based on big data, to individual farmers based on their circumstances, or to decision makers based on a wider geographic area to enable them to make more data-driven decisions.
- **Early warning—**analyzing data in order to, for example, identify disease or pest outbreaks before they spread.
- **Forecasting—**using big data tools to predict future trends, such as prices for specific crops in specific areas.
- **Financial services—**using big data to overcome barriers to providing credit and insurance to people who lack access to such financial services (for instance, small-scale producers with no credit history and limited collateral). Big data could potentially transform credit and insurance models by drawing in many more data sources about producers, their farm or other enterprise, the climate, and other factors. For example, GroVentures has aggregated and analyzed dispersed data sets from several dozen countries, enabling businesses to evaluate risk more effectively, resulting in more affordable crop insurance (World Bank 2015).

22 This figure originally appeared in the report "Personal Data: The Emergence of a New Asset Class," which was prepared by the World Economic Forum and Bain & Company in January 2011.

Some of these applications of big data will benefit producers directly, while others have the potential to help researchers and policy makers form recommendations and policies that will shape the future of farming in the face of climate change, increasing population, yield lags, demographic shifts, and depleting natural resources, among other challenges. Box 15.8 describes studies of prospective applications for big data in Senegalese agriculture.

Making sense of the data for these purposes requires specific analytical tools and methods. The five most common types of big data analytics are[23]:

- **Descriptive** analytics, the most common and widely used form of analytics, tells what has happened in the past and what is happening now. An example in agriculture is Esoko's MarketPlace service, which provides historical price trends of crops.
- **Diagnostic** analytics help explain why something has happened. This form of analytics includes the analysis of correlations and relationships in data to better understand causation. For example, Grameen Foundation's Community Knowledge Worker program, in partnership with Palantir, developed a platform that enables them, among other things, to understand the link between the application of good agronomic practices (GAP) and farm productivity (World Bank 2015).
- **Predictive** analytics uses predictive modeling to anticipate what will happen next based on past and current data. Back in 2007 agricultural consultancy firm Lanworth, now part of Thomson Reuters, was able to predict the volume of the U.S. corn crop with relative accuracy, using a mix of data including satellite images, weather forecasts, soil maps, crop conditions, and rotation patterns. In contrast, the U.S. Department of Agriculture, which used old-fashioned farmer surveys, overestimated that year's crop (Paynter 2008).
- **Prescriptive** analytics takes the trends identified in predictive analytics and recommends potential courses of action and their likely outcomes. It uses simulations, localized rules, and decision logic to identify options. For example, in Colombia, the International Center for Tropical Agriculture (CIAT) and the Colombian Rice Growers Federation (FEDEARROZ) developed a computer model including an artificial neural network[24] that incorporated 10 years of agricultural data, seasonal forecasts, and climate data. They predicted that a drought would occur and advised farmers against planting crops, saving those who adhered to their advice US$3.8 million (Clark 2014). The initiative received a UN Big Data Climate Challenge award and has plans to scale up over the next one to three years to include other crops and to expand into other countries in Latin America and Africa.
- **Cognitive** analytics uses a mix of artificial intelligence, machine learning algorithms, and in some case natural language processing to, in essence, mimic the cognitive capacity of humans. Although in many ways experimental and not yet widely available, cognitive analytics has the potential to completely change our ability to make sense of massive amounts of data in ways that our unaided minds are simply not capable of handling (Ronanki and Steier 2014).

The best way to measure the effectiveness of big data analytics in agriculture may be to determine how the insights revealed by the service have contributed to changes in efficiency and in

BOX 15.8. Big Data for Agriculture in Action

In 2014, Sonatel and Orange Group made anonymized call detail records of mobile subscribers in Senegal available to researchers as part of a Data for Development Challenge. Agriculture, one of the five priority areas, was the subject of four research papers analyzing whether mobile network data could be used for:

- Developing mobility profiles and calendars for food security and livelihood analysis.
- Understanding the genesis of millet prices in Senegal (including the roles of production, markets, and their failures).
- Improving disaster resilience through a visual analysis of call data records.
- Unraveling correlations between agricultural events and phone traffic.

The papers are available in the *Data for Development Challenge Senegal Book of Abstracts: Scientific Papers,* available at http://www.d4d.orange.com/fr/content/download/43453/406503/version/1/file/D4DChallengeSenegal_Book_of_Abstracts_Scientific_Papers.pdf.

Source: Authors.

23 The five types of analytics cited here are based on IBM's classifications (see http://www.ibm.com/analytics/us/en/analytics-technology/).

24 An artificial neural network is a type of machine learning algorithm that mirrors the interconnectedness of neurons in the human brain. These networks, which are helpful for analyzing complex and often incomplete data, are used for such things as detecting potential credit card fraud or the presence of explosives in airports (Kay 2001).

the cost of the overall outcomes that the service was intended to deliver. The capacity for realizing any benefit from big data is predicated on two factors at the output level. The first is that the analytic tools being used actually generate insights based on the data available, so (at a very basic level) one output that can be tracked is the number of insights generated. The second, and more important, output is the number of insights generated that are deemed to be accurate or worth acting upon.

The next section of this note highlights lessons and issues surrounding big data, and the three innovative practice summaries that follow present more detailed examples of how big data analytics is coming into its own in agriculture. The first two summaries describe open access resources: HarvestChoice, with spatially explicit, harmonized data layers on numerous major indicators for Sub-Saharan Africa, and the Crop Composition Database (CCDB), with rigorously vetted data on the nutritional composition of specific crop species. The third summary describes aWhere, a service that collects global meteorological data and matches them with information from farmers; the resulting data can be analyzed not only to develop personalized agronomic recommendations for farmers but also to contribute significantly to development policy, especially with respect to climate change.

LESSONS LEARNED

The case for using big data for analytics is pretty clear. The world is increasingly interconnected, volatile, and complex. The sheer volume of data that the world produces makes it simply impossible for humans to make sense of it all and react in real time. An added consideration is the inherent cognitive biases in human brains, which can lead to illogical conclusions or associations (Kahneman 2011).

Given these trends and concerns, it is not hard to see why such excitement surrounds the potential for big data to play a major role in contributing to the agricultural productivity gains that are needed to meet the world's food needs by 2050. Of course, realizing this potential is not as easy as simply turning on a switch. To maximize the impact of big data on agriculture, policymakers and practitioners will need to consider and address a number of issues, outlined here. In broad terms, these issues can be clustered into three groups: issues related to data ownership, access, and quality; issues related to analysis and interpretation; and issues related to implementation capacity.

Data Ownership, Access, and Quality

Determining who owns the collected data and how those data sets are protected are issues worthy of serious consideration. They have always been present to some degree in development—data collection is nothing new. With the advent of big data, what has changed is the role of third parties. Given the complexity of the computer systems needed for big data analytics, third parties increasingly store and analyze the data that organizations collect. Policymakers and practitioners need to think seriously about protections for these data, particularly data related to individuals, and implement them.

One example (mentioned earlier) of the risks involved in failing to protect data is the potential for individuals to be displaced from their land. In a context where land grabs are prevalent and rights to land are protected ineffectively, failure to prevent detailed information about the quality of farmers' soil and production capacity from falling into the wrong hands could put producers with high-quality land at greater risk. Anecdotal reports note that farmers in high-income countries have already expressed concern that they might be penalized if the government or environmental activists discover that they have applied fertilizer incorrectly (Gilpin 2014) or that traders might manipulate market prices based on access to information about what farmers are planting (Banham 2014). If producers perceive that their data are not secure or might be used against them, they may resist using tools that ultimately could help them.

Crowdsourced data are particularly affected by the potential of people to report incorrect information based on their perceptions rather than fact. It turns out, for example, that Google Flu Trends was not actually that good at predicting the occurrence of influenza; instead, it reflected the incidence of illnesses with flu-like symptoms (Fung 2014). If this tool had been applied to agriculture and used, for instance, to predict outbreaks of avian influenza, policy and commercial decisions based on such predictions could have mobilized resources to cope with an outbreak that did not exist.

Limitations in data sharing, compatibility, and availability also constitute barriers to fully realizing the potential of big data in agriculture. In a number of countries, for example, censuses and farm surveys are sporadic, incomplete, and often exclude smallholder farmers. The FAO's Global Strategy to Improve Agricultural and Rural Statistics is attempting to address this challenge through the introduction of an integrated survey system that countries can use to regularly collect and produce comprehensive agricultural data (Graziano da Silva 2016).

For the most part, structures and standards for sharing proprietary data are also lacking. Unlocking those data will require public-private partnerships; it will also require stakeholders to understand how they will benefit, both individually and collectively, from doing so (World Economic Forum 2015).

A strong open data movement, which the World Bank and others are promoting, seeks to bring more of the world's data into the public domain, including agricultural data (see box 15.9 and the innovative practice summary below). A vast share of these data remain closed to the public, however. Of the 1,290 data sets of public records surveyed from 86 countries by the Open Data Barometer in 2014, only slightly more than 10 percent met the definition of open data by being "published in bulk, machine-readable formats, and under an open license" (World Wide Web Foundation 2015).

Even members of the development community, which should have many incentives to share data with each other, have responded unevenly to the call for open data. Part of the challenge in making data public is to develop internal processes for managing data and to compile data sets that may be sitting on hundreds of different hard drives around the organization. Another part of the challenge is to overcome resistance to change among individuals who are simply used to handling their data in one way, and have yet to adjust to new data management processes. Both of these issues speak to the need for organizations in the development community to put forward very clear guidance on open data set requirements and to implement clear data management processes that are effectively communicated to all employees.

Not surprisingly, gaining access to data in the private domain is even more complicated. Many private businesses are reluctant to share data with third parties without either selling the data to them and/or requiring legal agreements, such as nondisclosure agreements. CDRs are a good example. CDRs collected by mobile network operators indicate the approximate location of all calls made on their network. The analysis of CDRs has been suggested as a way to track population movements during times of conflict or disease outbreaks, but the network operators' limited commercial interest in sharing such data is often overlooked. The analysis of CDRs also generally ignores principles of informed consent, since individual callers have not consented to having their data analyzed for those purposes (Letouzé and Vinck 2014).

BOX 15.9. Suggestions for Unleashing the Power of Data for Agriculture

Participants in the 2015 International Open Data Conference (Canada) identified seven ways that data can be made more effective in agriculture: (1) open up data, (2) identify data users, (3) bring intermediaries into the game, (4) develop new tools for data collection, (5) look beyond technology, (6) foster cross-sector collaboration, and (7) address the need for disaggregated data. More detailed information can be found online here.[a]

Source: Authors.
a. Halais 2015 (https://www.devex.com/news/7-challenges-the-agriculture-sector-must -address-to-unleash-its-data-revolution-86310).

Analysis and Interpretation

The challenges related to analysis and interpretation can be broadly classified into three types of limitations: data, human, and machine. Data limitations refer to shortcomings in data sets that prevent their meaningful analysis, such as an insufficient amount of data, erroneous data, or unreadable data.

Many of the human limitations relate to cognitive biases and poor interpretation of the data or of their implications.[25] Even in the case of cognitive analytics, which aims to reduce the potential effect of such limitations, ultimately it is still humans who will make decisions based on the outputs from the analytic tools. At least in the near future, human judgment will retain a significant role in benefiting from analytic insights, because most big data analytics will remain outside the realm of cognitive analytics.

The supply/demand gap between the number of professionals with the requisite skills to analyze and make decisions from big data, and the number of professionals needed across all industries with an interest in big data, has been reported by a number of sources over the past few years. Governments, academic institutions, and development organizations, in addition to the private sector, need to address this talent gap in the agricultural sector through capacity building and education. Skills must be developed at the farm level as well as among development practitioners and researchers, all of whom will benefit from better understanding how to make sense of data and act upon that knowledge.

Although, one day, artificial intelligence, driven by machine learning algorithms, may be virtually indistinguishable from human intelligence, machines—at least for the present—also have limitations. Even Watson, IBM's supercomputer—a system much more powerful by far than anything the reader is likely to access in the near future—made very basic mistakes in common sense during its triumphant run on the U.S. television show *Jeopardy!* several years back (Hamm 2011).

25 The UN Global Pulse's Big Data for Development: Challenges & Opportunities has a thorough section on challenges in the analysis of big data, mostly related to human limitations.

Implementation Capacity

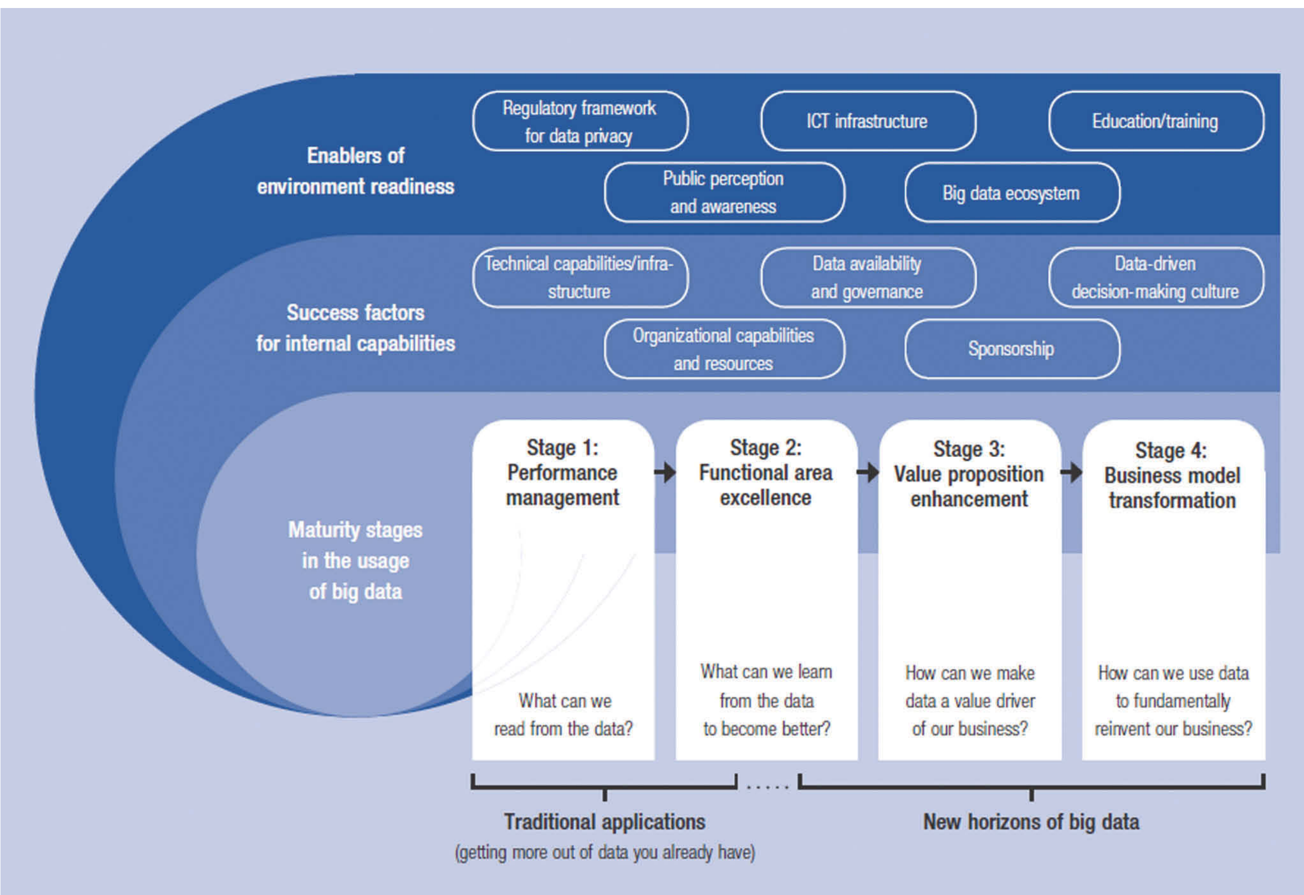
A 2013 worldwide survey by Gartner found that 64 percent of organizations had invested in or were planning to invest in big data systems (Gartner 2013). Over the next few years, this figure is likely to continue to grow rapidly, as an increasing number of medium-sized and large agribusinesses in high- to low-income countries will probably invest in some form of big data solution. A good portion of those firms will also probably fall victim to at least one of the eight implementation challenges identified by Svetlana Sicular at Gartner (Sicular 2014): (1) management inertia, (2) selecting wrong use cases, (3) asking wrong questions, (4) lacking the right skills, (5) unanticipated problems that are wider than just a big data technology, (6) disagreement on the enterprise strategy, (7) siloed big data, and (8) solution avoidance.

For governments seeking to implement big data systems to support agricultural development, the challenges are likely to be even greater, given the scale. In an examination of the sources of the initial failure of the U.S. healthcare.gov website, Clifford Winston of the Brookings Institution identified four primary contributors: (1) limited technical expertise and an overreliance on contractors; (2) little, if any, rigorous and transparent ongoing assessment because of a fear of exposing problems; (3) a status-quo bias and an inflexibility and inability to make important changes in managing a project; and (4) constraints that may affect budgeting and adoption of state-of-the-art technology (Winston 2013).

All such factors should be considered and taken into account in supporting and deploying any big data system. The relevance of each factor depends somewhat on how much the use of big data has advanced in a particular country or organization. Figure 15.3 provides a useful framework for understanding the different stages of big data use in an organization, as well as the internal capabilities and ecosystem enablers that need to be in place first.

26 This figure originally appeared in the chapter "Big Data Maturity: An Action Plan for Policymakers and Executives" within The Global Information Technology Report 2014.

FIGURE 15.3. Big Data Maturity Framework



Source: Strategy& (formerly Booz & Company), part of the PwC network of firms.[26]
Note: Used with permission. New request for permission is required if image is reused.

Helpful resources

- Big Data in Action for Development, http://live.worldbank.org/big-data-in-action-for-development.
- Big Data, Big Impact: New Possibilities for International Development, https://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development.
- Big Data for Development: Challenges & Opportunities, http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf.

INNOVATIVE PRACTICE SUMMARY
Generating Open Access, Spatially Explicit Data Sets, and Analyses for More Productive Farming and Better Livelihoods in Africa

A growing community of scientists is taking advantage of publicly available global data sets remote sensing imagery, GIS, computer modeling, and georeferenced data pooled from the bottom up (or crowdsourced) to capture the heterogeneity of humans and agriculture and gain a better understanding of the complex spatial relationships between agriculture, the environment, climate change, and social well-being.[27] A problem in many settings, but especially in low-income countries and among low-income and vulnerable populations, is that critical data are difficult to obtain from administrative offices or to generate through experiments and observations on the ground. The growing availability of cross-harmonized data and geospatial tools is helping to alleviate some of those constraints.

HarvestChoice is hub to a large number (more than 750 and growing) of spatially explicit, harmonized data layers for Sub-Saharan Africa, including indicators of health and poverty variables, agricultural production and area, climate and soil, and access to markets (box 15.10). Web-based data analytics tools, such as HarvestChoice's Mappr, allow development practitioners and analysts to dip into HarvestChoice's core data holdings and visualize geographical impacts for investment and policy design at scale (HarvestChoice 2012). More advanced users can plug directly into HarvestChoice data through an open API[28] (HarvestChoice 2014b).

27 This innovative practice summary was developed by Cindy Cox and Jawoo Koo at the International Food Policy Research Institute (IFPRI).

28 A publicly available application programming interface.

Data are harmonized across domains and country borders, allowing for complex spatial analyses and evidence-based investment strategies. For example, HarvestChoice's Spatial Production Allocation Model (SPAM) uses a cross-entropy approach on a variety of inputs, from subnational crop production statistics to market information, to generate plausible, disaggregated estimates of crop distribution for 42 crops and their performance around the globe (You et al. 2014). SPAM results are useful for understanding production and land-use patterns and for identifying geographical trends. They also provide a means for understanding the causalities of cropping outcomes within disaggregated units. SPAM data can be manipulated via Mappr or downloaded from www.mapspam.info.

Pooled data from georeferenced household surveys, such as Demographic and Health Surveys[29] (DHS) and the LSMS[30], are filling critical information gaps with respect to nutrition, health, gender-related variables, wealth (consumer durables, housing characteristics), education, and access to services (water, sanitation, health facilities, schools). When it comes to agricultural activities, income, and infrastructure, however, the survey data are weak (although improving). This is where the interoperability of harmonized data becomes important. By combining population data from household surveys with HarvestChoice spatial data on agriculture, biophysical characteristics, and market access, it is possible to produce a well-rounded set of variables and facilitate studies on nutrition and agriculture across or within countries.

HarvestChoice currently holds many spatial layers on nutrition and dietary outcomes based on crowdsourced surveys, allowing Mappr users to visualize the spatial distribution of diet quality and nutritional outcomes across subnational regions in Sub-Saharan Africa. Such data can support analyses of how diet and nutrition are related to market characteristics, the environment, and agricultural systems, and they can provide the context for understanding the scalability of research outcomes.

Economic and crop modelers are also increasingly taking advantage of granular data sets, especially those harmonized on high-resolution global grids in modeling analyses that explore future consequences of climate change (Nelson 2014; Rosegrant 2014). Crop models such as Decision Support System for Agrotechnology Transfer (DSSAT) make

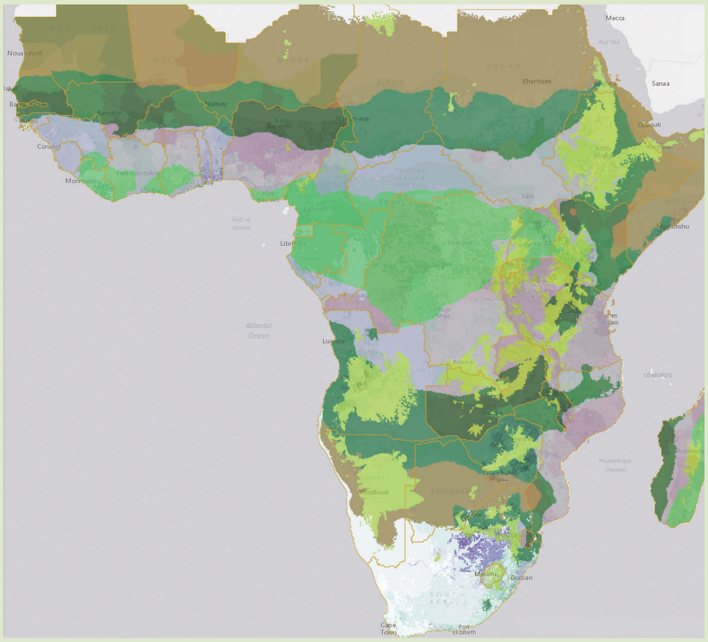
29 Available at http://dhsprogram.com/data.

30 Available at http://go.worldbank.org/BCLXW38HY0.

BOX 15.10. Aggregating and Visualizing Data in Mappr

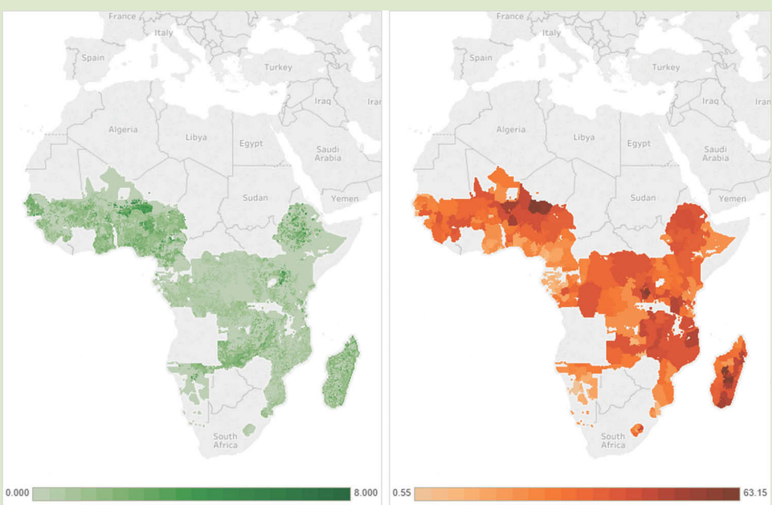
Mappr users can aggregate ~10 × 10 kilometer pixels in meaningful ways, such as by farming system, watershed, or agro-ecological zone. HarvestChoice uses remote sensing, GIS, open source subnational data sets, crop and economic modeling (DSSAT and IMPACT), and georeferenced household surveys to generate over 750 layers of subnational socio-economic and biophysical data for Sub-Saharan Africa; see, for example, images B15.10.1 and B15.10.2.

IMAGE B15.10.1. HarvestChoice's Mappr



HarvestChoice's Mappr (http://apps.harvestchoice.org/mappr) allows users to easily explore +700 multi-disciplinary geospatial indicators across SSA without needing to use advanced GIS software or spatial analysis skills. Users can browse the data catalog, select multiple indicators of interest, visualize them on the map, and execute a set of spatial analysis. This screenshot shows an example of spatial domain analysis output generated from three user-selected indicators (rural poverty, maize harvest area, and growing period) presented on the map and a series of charts.

IMAGE B15.10.2. HarvestChoice's grid-based multi-disciplinary indicator database (CELL5M)



Maize yield (kg/ha, 2005)

Stunting, Moderate (z-score –6 to –2)

Learn more about CELL5M, which powers a suite of spatial targeting analyses and applications, at http://dx.doi.org/10.12688/f1000research.9682.1. This tool (available online at https://public.tableau.com/profile/ifpri.td.hc#!/vizhome/cell5m_a4nh_v2_ssa/CELL5M_A4NH) provides an easy-to-use, interactive indicator-level mapping and filtering interface to identify the areas meeting multiple search criteria of agriculture and nutrition baseline.

More than just pretty maps, georeferenced data can help development practitioners visualize where populations are most vulnerable, the farming systems they most depend on, the biological and geophysical constraints and risks that limit farm productivity, the investments and innovations that could raise farm productivity most sustainably, and the broader impacts of such change.

Source: Authors.
Note: Images used with permission. New request for permission is required if image is reused.

it possible to explore global changes in agricultural productivity (Jones et al. 2003). Results from crop models can be integrated with economic models (such as the IMPACT and DREAM models) to study alternative projections of global food supply, demand, trade, prices, and food security (HarvestChoice 1995; Rosegrant 2008). Agritech Toolbox, a geo-tool available from HarvestChoice, allows users to explore simulation results from those two types of models and visualize the impact of agricultural technologies around the globe (HarvestChoice 2014a). This analytical capacity is particularly important in regions of the world where the effects of global changes in the environment, including the effects of climate change, are most pressing and consequential.

INNOVATIVE PRACTICE SUMMARY
The ILSI Crop Composition Database

The CCDB (https://cropcomposition.org), launched in 2003, is a curated, open resource that provides data on the natural variability in the nutritional composition of specific crop species (e.g., information on nutrients, anti-nutrients, and secondary metabolites) (Alba et al. 2010; Ridley et al. 2004).[31] These data have multiple uses, although the CCDB was originally developed to provide information for risk assessors and regulators to undertake intraspecies comparative assessments of the nutritional content of conventional versus transgenic crops (CAC 2003). The non-profit International Life Sciences Institute (ILSI) Research Foundation maintains the CCDB, and criteria for accepting data are overseen by the CCDB Working Group, which comprises scientists from the public and private sectors. The most recent version of the CCDB was released in 2014 and includes more than 840,000 data points representing 3,150 compositional components.

Data in the CCDB are derived from numerous samples of hybrids and/or varieties cultivated in controlled field trials using standard commercial cultivation practices at various locations throughout the world. Representative plant samples are obtained from field-grown crops with known production locations and dates. The analytical methods used to generate the data must be indicated, validated, and use certified or historically verified standards. Data are uploaded in a standardized format by an authorized provider using a secure data provider tool. As the comma-delimited file is uploaded, it is checked for format and duplication at the file and sample level. The content of successfully uploaded data is then tested for validity and consistency. Users can query the database to generate mean levels and ranges of nutritional components in various crop species. Environmental factors such as soil type and temperature can affect the levels of important nutrients in plants, and the moisture content can vary based on field conditions at harvest and when samples are handled. The database includes features that allow the user to retrieve a subset of data for samples produced in a specific year or location, and the analyst search filter can be applied to retrieve a predetermined subset of data.

The CCDB is accessible to scientists from academia, government agencies, and industry as well as to the general public, and it is a well-used resource. From July 2014 to July 2015, the CCDB logged 81,838 unique site visits from users in 122 countries around the world. It is referenced in peer-reviewed publications, regulatory guidance documents, and in many regulatory dossiers submitted in support of genetically engineered food safety assessments. This database complements existing food and nutrient databases, and it is an important but probably underused resource for food scientists, nutrition practitioners, and others interested in the interface between agriculture and nutrition.

One of the strengths of the CCDB—the completeness and quality of the data sets for each of the subject crops—is also potentially one of its limitations, however. The analytical rigor required for data submitted to the CCDB means that sample testing is expensive, so it is not surprising that most data has been provided (at no charge) by the private sector, and for a very limited range of crops (currently canola, field corn, sweet corn, cotton, rice, and soybeans). The ILSI Research Foundation and the CCDB Working Group are committed to including data for other crop species, particularly of important staple foods. For such data to become available, public sector breeding programs, as well as breeding programs run by small and medium-sized private firms, must be able to submit data, but it is also imperative to ensure that data for new crops are verifiable and robust. Resolving how to balance these imperatives remains a significant challenge.

INNOVATIVE PRACTICE SUMMARY
Using Big Data to Provide Localized Weather and Agronomic Information to Producers

Throughout the world, a great number of producers use traditional knowledge of weather-related signs and conditions to make their agricultural choices. This knowledge—which has sustained countless generations of

31 This innovative practice summary was written by Morven McLean of the ILSI Research Foundation.

producers—has become a less reliable guide as more variable weather patterns have brought less predictability, higher risk, and a growing sense of uncertainty to farmers and to agriculture.[32]

Uncertainty can affect behavior by making people even more averse to risk. Farmers who are more risk averse may be less likely to experiment with or adopt new approaches, including approaches that could increase agricultural and environmental sustainability in an era of climate change and less predictable weather. The connection between uncertainty and subsequent environmental degradation can create a vicious cycle in which the impact of increased weather variability on agriculture continues to grow. Services that tailor agricultural information to highly local settings can help farmers adjust to weather variability, increase the productivity and profitability of farming, and create an opportunity to improve sustainable food production.

By combining localized weather information with farmer-specific tips, aWhere is seeking to address these issues. Its predictive analytic platform collects over 1 billion data points daily on temperature, rainfall, humidity, solar radiation, and wind from satellites, weather aggregators, and drone operators, resulting in a global meteorological data set that covers all agricultural geographies. As a result, all weather data, from a 20-year history to 8-day forecasts, are consistently available globally. At the same time, the platform is quite localized, matching farmer-specific agricultural tips with growth stage models that are specific to each region.

While aWhere provides historical weather information, forecasts, and agronomic models, the platform relies on data inputs from partners for other types of information, such as cropping calendars and agronomic tips for each crop and variety. This information is generally acquired through local resources such as agricultural extension services, universities, and local knowledge. The combination of aWhere's agronomic models and a well-defined crop calendar effectively results in a personalized crop calendar for every farmer.

The relevance of the intelligence farmers receive from aWhere depends, in turn, on how much information they provide to the service. If the farmer provides the latitude and longitude of his or her farm and the crops planted, the information can be more tailored to that farmer's circumstances. In practice, aWhere has found that farmers typically provide the name of a village, for which the coordinates can be identified. Given that many farmers do not have GPS-enabled phones or, if they do, may not know how to collect their coordinates, it may be necessary for someone else to collect that information on their behalf. Once their specific information is input into the platform, however, they can begin to receive crop-specific advice, localized weather forecasts, details on nearby input providers, and local market prices.

Key Lessons from aWhere

- Combining data from multiple sources provides a longitudinal view of climate effects and produces insights that can be extremely beneficial to farmers.
- Simply collecting the data and generating insights is not enough; the use of technology to make sure that the message reaches the right stakeholders at the right time is also important.

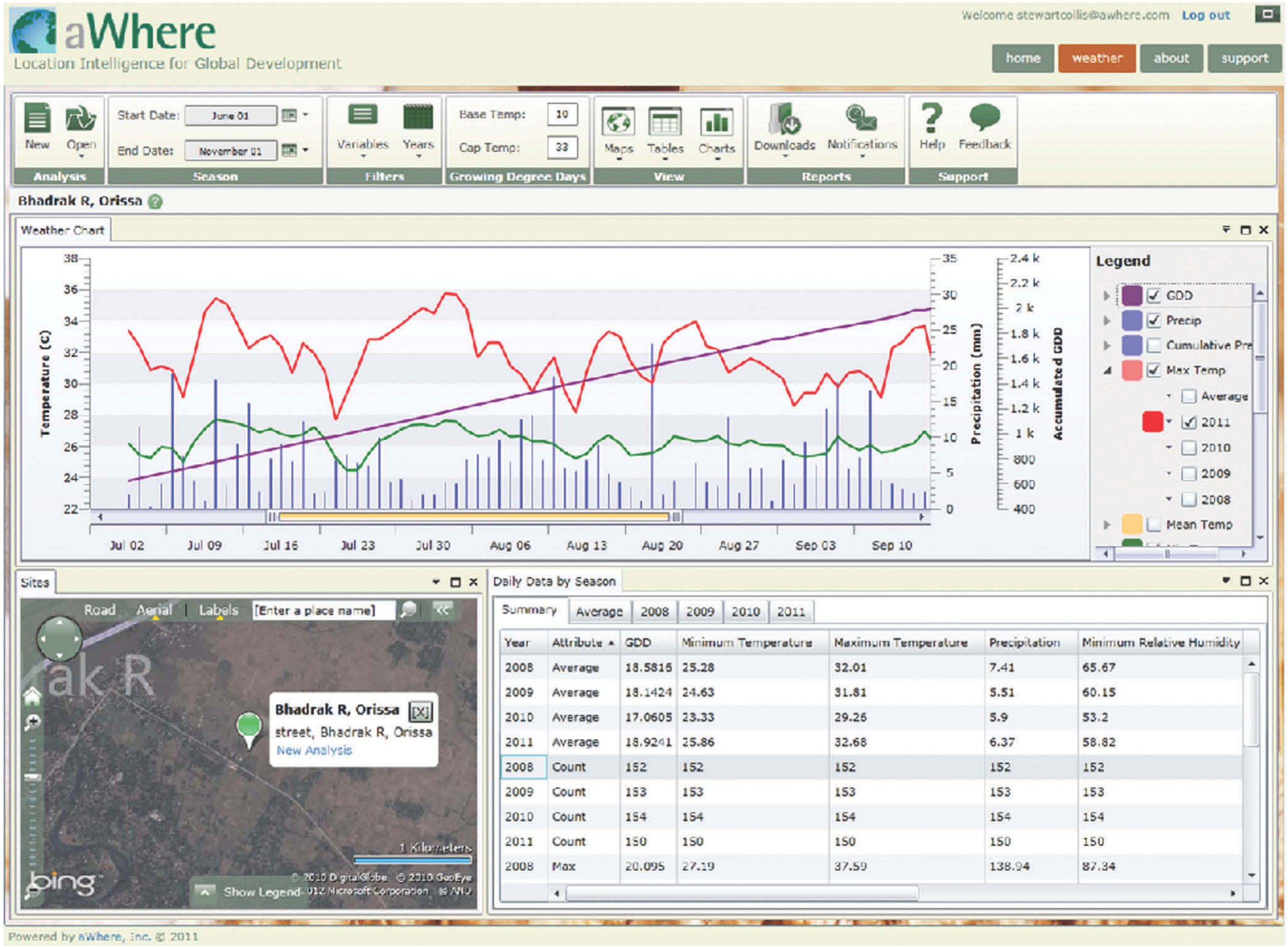
Most of a Where clients are based in the United States, although the service is expanding into Africa, Asia, and the Caribbean, where it partners with agricultural information providers such as Esoko in Ghana, which makes aWhere's weather data and agronomic models available to its clients (Storum 2015).

Access to this combination of weather intelligence and agronomic recommendations is expected to help farmers make more informed decisions—for example, to delay planting because of a projected delay in the start of the seasonal rains. Policy makers, researchers, and development practitioners can also use aWhere's online platform for their own analysis and decision making (image 15.4). By importing their own data sets into the platform, they can identify correlations that might be helpful, such as correlations between rainfall levels and market prices or disease outbreaks.

An added capability is that researchers and commercial practitioners can combine weather data with historical information on crop yields to generate field-specific agronomic models, as well as management recommendations for weather-smart agriculture. These models and recommendations generate new agricultural intelligence that can enhance traditional agricultural practices and provide guidance to farmers for mitigating the risks of adverse weather events and climate variability. This new, real-time information

32 The majority of the text provided in this summary was adapted from content that came directly from Tarah Speck at aWhere. Some of the text is from the module's primary author, and was drawn from publicly available information about aWhere. To learn more about aWhere, visit http://www.awhere.com/.

IMAGE 15.4. Screenshot of aWhere's Online Dashboard



Source: aWhere.
Note: Used with permission. New request for permission is required if image is reused.

allows farmers and stakeholders across the world to make evidence-based agricultural decisions and optimize farming practices as the dynamics of agriculture change.

All of aWhere's weather and agronomic data can be delivered through RESTful APIs, allowing for their integration into customized apps or widgets. More details, including samples and JavaScript visualizations, can be found on their developer's portal (http://developer.awhere.com/).

REFERENCES

Alba, R., A. Phillips, S. Mackie, N. Gillikin, C. Maxwell, P. Brune, W. Ridley, J. Fitzpatrick, M. Levine, and S. Harris. 2010. "Improvements to the International Life Sciences Institute Crop Composition Database." *Journal of Food Composition and Analysis* 23: 741–48.

Banham, Russ. 2014. "Who Owns Farmers' Big Data?" *Forbes Brand Voice*. http://www.forbes.com/sites/emc/2014/07/08/who-owns-farmers-big-data/.

Bilbao-Osorio, Beñat. 2014. *The Global Information Technology Report 2014: Rewards and Risks of Big Data*. World Economic Forum, Cologny. http://www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf.

BIS Research. 2014. *Global Precision Agriculture Market Analysis & Forecast (2015–2022) Technology (VRA, Soil Mapping, Yield Monitoring, Precision Irrigation, Others), Components and Systems*. http://www.researchandmarkets.com/research/fk898n/global_precision.

Bujoreanu, Luda. 2013. "The Power of Mobile: Saving Uganda's Banana Crop." *Information and Communications for Development* (blog). http://blogs.worldbank.org/ic4d/the-power-of-mobile-saving-ugandas-banana-crop.

Burns, Matt. 2015. "3D Robotics Taps Qualcomm for $50M Series C and Mobile Tech." *TechCrunch*. http://techcrunch.com/2015/02/26/3d-robotics-taps-qualcomm-for-50m-series-c-and-mobile-tech/.

CAC (Codex Alimentarius Commission). 2003. "Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant-DNA Plants." CAC/GL 45-2003, Rome.

Clark, Helen. 2014. "How Big Data Is Helping Farmers Save Millions." *GizMag*. http://www.gizmag.com/big-data-crops-climate-change/34400/.

Conor, Anthony. 2014. "Benefits and Limitations of Crowdsourcing Agricultural Data." *AGROAM* (blog). http://www.agroam.org/blog/benefits-and-limitations-of-crowdsourcing-agricultural-data/.

Crawford, John, John Hutchins, Richard Tiffin, and Alistair Stott. 2015. "AIMS: Putting the UK at Heart of the Big Data Revolution in Agriculture." *AgriTech Strategy* (blog). https://agritech.blog.gov.uk/2015/03/27/aims-putting-the-uk-at-heart-of-the-big-data-revolution-in-agriculture/.

Crimson Hexagon. 2011. "Twitter and Perceptions of Crisis Related Stress." UN Global Pulse. http://www.unglobalpulse.org/projects/twitter-and-perceptions-crisis-related-stress.

CTA. 2015. "Satellites and Mobile Phones Improve Crop Productivity in Sudan." http://www.cta.int/en/article/2015-06-17/satellites-and-mobile-phones-improve-crop-productivity-in-sudan.html.

Dawson, Ross, and Steve Bynghall. 2012. *Getting Results from Crowds: The Definitive Guide to Using Crowdsourcing to Grow Your Business*. 2nd ed. Essen: Advanced Human Technologies.

Essiet, Daniel. 2015. "Combating Challenges of Poor Data-driven Agriculture." *The Nation* (Nigeria). http://thenationonlineng.net/combating-challenges-of-poor-data-driven-agriculture/.

European Space Agency. 2014. "Mobile App Provides Weather Forecasts for Rwandan Coffee Farmers." *Phys.org*. http://phys.org/news/2014-03-mobile-app-weather-rwandan-coffee.html#jCp.

Fahlgren, Noah, Malia A. Gehan, and Ivan Baxter. 2015. "Lights, Camera, Action: High-Throughput Plant Phenotyping Is Ready for a Close-Up." *Current Opinion in Plant Biology* 24: 93–99. http://www.sciencedirect.com/science/article/pii/S1369526615000266.

Ferster, Warren. 2014. "U.S. Government Eases Restrictions on DigitalGlobe." *Space News*. http://spacenews.com/40874us-government-eases-restrictions-on-digitalglobe/.

Freischlad, Nadine. 2015. "Drones over the Rice Paddy: Ci-Agriculture Brings Smart Tech to the Field." *Tech in Asia*. https://www.techinasia.com/ci-agriculture-precision-farming-indonesia/.

Fung, Kaiser. 2014. "Google Flu Trends' Failure Shows Good Data > Big Data." *Harvard Business Review*. https://hbr.org/2014/03/google-flu-trends-failure-shows-good-data-big-data/.

Gartner. 2012. "Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data by 2015." http://www.gartner.com/newsroom/id/2207915.

———. 2013. "Gartner Survey Reveals that 64 Percent of Organizations Have Invested or Plan to Invest in Big Data in 2013." http://www.gartner.com/newsroom/id/2593815.

Gilpin, Lyndsey. 2014. "How Big Data Is going to Help Feed Nine Billion People by 2050." *TechRepublic*. http://www.techrepublic.com/article/how-big-data-is-going-to-help-feed-9-billion-people-by-2050/.

Graziano da Silva, José. 2016. "Big Data, Small Farms." *Foreign Affairs*. https://www.foreignaffairs.com/sponsored/big-data-small-farms.

Greatrex, Helen, James Hansen, Samantha Garvin, Rahel Diro, Sari Blakeley, Margot Le Guen, Kolli Rao, and Daniel Osgood. 2015. "Scaling Up Index Insurance for Smallholder Farmers: Recent Evidence and Insights." CCAFS Report 14, CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Copenhagen. https://cgspace.cgiar.org/bitstream/handle/10568/53101/CCAFS_Report14.pdf.

Gustafson, Thomas Andrew. 2013. "I'm Not Just Gaming, Ma! I'm Helping the World's Farmers." *NPR*. http://www.npr.org/sections/thesalt/2013/11/25/247210031/i-m-not-just-gaming-ma-i-m-helping-the-world-s-farmers.

Halais, Flavie. 2015. "7 Challenges the Agriculture Sector Must Address to Unleash its Data Revolution." *Devex*. https://www.devex.com/news/7-challenges-the-agriculture-sector-must-address-to-unleash-its-data-revolution-86310.

Hamm, Steve. 2011. "Watson on Jeopardy! Day One: Man vs. Machine for Global Bragging Rights." *A Smarter Planet* (blog). http://asmarterplanet.com/blog/2011/02/watson-on-jeopardy-day-one-man-vs-machine-for-global-bragging-rights.html.

HarvestChoice.1995. *Dynamic Research Evaluation for Management (DREAM) 3.1*. Washington, DC: International Food Policy Research Institute (IFPRI) and University of Minnesota. http://hdl.handle.net/1902.1/18230.

———. 2012. "Visualize HarvestChoice Indicators: MAPPR." International Food Policy Research Institute, Washington, DC, and University of Minnesota, St. Paul. http://harvestchoice.org/node/2258.

———. 2014a. "AgriTech Toolbox." International Food Policy Research Institute, Washington, DC, and University of Minnesota, St. Paul. http://harvestchoice.org/node/9034.

———. 2014b. "HarvestChoice Data Services." International Food Policy Research Institute, Washington, DC, and University of Minnesota, St. Paul. http://harvestchoice.org/node/9025.

Hosenally, Nawsheen. 2012. "Mapping Trees for Food Security." *ICT Update* 69. http://ictupdate.cta.int/Feature-Articles/Mapping-trees-for-food-security/

IDC. 2014. "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things." EMC². http://www.emc.com/infographics/digital-universe-2014.htm.

Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, and J. T. Ritchie. 2003. "The DSSAT Cropping System Model." *European Journal of Agronomy* 18 (3): 235–65.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.

Kay, Alexx. 2001. "Artificial Neural Networks." *Computerworld*. http://www.computerworld.com/article/2591759/app-development/artificial-neural-networks.html.

Kelly, Jeff. 2015. "Executive Summary: Big Data Vendor Revenue and Market Forecast, 2011–2026." *Wikibon*. http://premium.wikibon.com/executive-summary-big-data-vendor-revenue-and-market-forecast-2011-2026/.

Laney, Doug. 2001. "3D Data Management: Controlling Data Volume, Velocity, and Variety." *META Group* (blog). http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lavars, Nick. 2015. "MIT's Crop-Saving Drones at the Ready." *Gizmag*. http://www.gizmag.com/mit-crop-drones/37526/

Lawler, Ryan. 2015. "Planet Labs Nabs $95 Million and a New COO to Cover the Earth with Flocks of Tiny Satellites." *TechCrunch*. http://techcrunch.com/2015/01/20/planet-labs-95m/.

Leclerc, Rob, and Melissa Tilney. 2015. "AgTech Is the New Queen of Green." *TechCrunch.* http://techcrunch.com/2015/04/01/the-new-queen-of-green/.

Letouzé, Emmanuel, and Patrick Vinck. 2014. "The Politics and Ethics of CDR Analytics." Data-Pop Alliance White Paper Series. http://static1.squarespace.com/static/531a2b4be4b009ca7e474c05/t/54b97f82e4b0ff9569874fe9/1421442946517/WhitePaperCDRsEthicFrameworkDec10-2014Draft-2.pdf.

Marr, Bernard. 2015. "From Farming to Big Data: The Amazing Story of John Deere." *Data Science Central* (blog). http://www.datasciencecentral.com/profiles/blogs/from-farming-to-big-data-the-amazing-story-of-john-deere.

Martin, Steven W., James Hanks, Aubrey Harris, Gene Willis, and Swagata Banerjee. 2005. "Estimating Total Costs and Possible Returns from Precision Farming Practices." *Crop Management*. http://naldc.nal.usda.gov/download/11889/PDF.

McKinsey Global Institute. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

McLaren, Robin. 2015. *Managing Unintended Consequences of Democratizing Land Rights.* http://www.fig.net/resources/proceedings/fig_proceedings/fig2015/papers/ts02b/TS02B_mclaren_7687.pdf.

Nelson, G. C., H. Valinb, R. D. Sands, P. Havlík, H. Ahammad, D. Deryng, J. Elliott, S. Fujimori, T. Hasegawa, E. Heyhoe, P. Kyle, M. Von Lampe, H. Lotze-Campen, D. M. d'Croz, H. van Meijl, D. van der Mensbrugghe, C. Müllerk, A. Popp, R. Robertson, S. Robinson, E. Schmid, C. Schmitz, A. Tabeau, and D. Willenbockel. 2014. "Climate Change Effects on Agriculture: Economic Responses to Biophysical Shocks." *Proceedings of the National Academy of Sciences of the United States of America* 111 (9). http://dx.doi.org/10.1073/pnas.1222465110.

Paynter, Ben. 2008. "Feeding the Masses: Data In, Crop Predictions Out." *Wired Magazine*. http://archive.wired.com/science/discoveries/magazine/16-07/pb_feeding.

Practical Action. 2012. *Poor People's Energy Outlook 2012: Energy for Earning a Living*. Bourton, UK: Practical Action Publishing. http://practicalaction.org/ppeo2012.

Quinn, Sara. 2014. "Creating a Community of Practice in Sub-Saharan Africa to Utilize Unmanned Aerial Vehicle-based Tools for Agricultural Development." International Potato Center. http://cipotato.org/press-room/blogs/creating-practice-sub-saharan-africa-unmanned-aerial-vehicle-based-tools-agricultural-development/.

Reader, Ruth. 2015. "Google Ventures Leads $15M Investment in Big Data for Farmers." *VentureBeat*. http://venturebeat.com/2015/05/19/google-ventures-leads-15m-investment-in-big-data-for-farmers/.

Ridley, W. P., Shillito, R. D., Coats, I., Steiner, H-Y, Shawgo, M., Phillips, A., Dussold, P., Kurtyka, L., 2004. "Development of the International Life Sciences Institute Crop Composition Database." *Journal of Food Composition and Analysis* 17: 423–38.

Ronanki, Rajeev, and David Steier. 2014. "Cognitive Analytics." Tech Trends 2014. http://dupress.com/articles/2014-tech-trends-cognitive-analytics/.

Rosegrant, M. W., Msangi, S., Ringler, C., Sulser, T. B., Zhu, T., and Cline, S. A. 2008. "International Model for Policy Analysis of Agricultural Commodities and Trade (IMPACT): Model Description." 42, International Food Policy Research Institute, Washington, DC.

Rosegrant, M. W., J. Koo, N. Cenacchi, C. Ringler, R. D. Robertson, M. Fisher, C. M. Cox, K. Garret, N. D. Perez, and P. Sabbagh. 2014. *Food Security in a World of Growing Natural Resource Scarcity: The Role of Agricultural Technologies*. Washington, DC: International Food Policy Research Institute. http://dx.doi.org/10.2499/9780896298477.

Saporito, Patricia. 2014. "2 More Big Data V's—Value and Veracity." *SAP Business Innovations* (blog). http://blogs.sap.com/innovation/big-data/2-more-big-data-vs-value-and-veracity-01242817.

See, Linda. 2013. "Citizen Scientists Rival Experts in Analyzing Land-Cover Data." http://www.iiasa.ac.at/web/home/about/news/PLOS_ONE__Citizen_scientists_rival_experts_in_analyz.en.html.

See, Linda, I. McCallum, S. Fritz, C. Perger, F. Kraxner, M. Obersteiner, U. D. Baruah, N. Mili, and N. R. Kalitas. 2013. "Mapping Cropland in Ethiopia Using Crowdsourcing." *International Journal of Geosciences* 4: 6–13. http://dx.doi.org/10.4236/ijg.2013.46A1002.

Sicular, Svetlana. 2014. "Big Botched Data." *Gartner Blog Network*. http://blogs.gartner.com/svetlana-sicular/big-botched-data/.

Silversmith, Alex, and Drew Tulchin. 2013. "Crowdsourcing Applications for Agricultural Development in Africa." USAID, Washington, DC. http://pdf.usaid.gov/pdf_docs/pa00j7p7.pdf.

Storum, Doug. 2015. "Farmers Stay Aware with aWhere's Weather Data." *Boulderopolis*. http://boulderopolis.com/farmers-stay-aware-with-awheres-weather-data/.

Thomasson, Alex. 2015. "The Future of Farming: Drones, Robots, and GPS." *The Conversation*. http://mashable.com/2015/03/22/farmers-drones-robots/.

Tsotsis, Alexia. 2013. "Monsanto Buys Weather Big Data Company Climate Corporation for Around $1.1B." *TechCrunch*. http://techcrunch.com/2013/10/02/monsanto-acquires-weather-big-data-company-climate-corporation-for-930m/.

United Nations, Department of Economic and Social Affairs, Population Division. (2015). *World Population Prospects: The 2015 Revision, Key Findings and Advance Tables*. ESA/WP.241. New York: United Nations.

USDA (U.S. Department of Agriculture). 2005. "Agricultural Aircraft Offer a Different View of Remote Sensing." *AgResearch Magazine.* http://agresearchmag.ars.usda.gov/2005/mar/remote.

Winston, Clifford. 2013. "Government Implementation of Large-Scale Projects: Government Failure, Its Sources, and Implications for the ACA Website Launch." Brookings Institution. http://www.brookings.edu/research/testimony/2013/12/04-government-implementation-large-scale-projects-winston

World Bank. 2014. "Big Data in Action for Development." http://live.worldbank.org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version.pdf.

———. 2015. Big Data Solutions: Innovative Approaches to Overcoming Agricultural Challenges in Developing Nations by Harnessing the Power of Analytics. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/26436.

World Economic Forum. 2011. "Personal Data: The Emergence of a New Asset Class." http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.

———. 2015. "Data-Driven Development Pathways for Progress." http://www3.weforum.org/docs/WEFUSA_DataDrivenDevelopment_Report2015.pdf.

World Wide Web Foundation. 2015. *Open Data Barometer Global Report*. 2nd ed. http://barometer.opendataresearch.org/assets/downloads/Open%20Data%20Barometer%20-%20Global%20Report%20-%202nd%20Edition%20-%20PRINT.pdf.

You, L., S. Wood, U. Wood-Sichra, and W. Wu. 2014. "Generating Global Crop Distribution Maps: From Census to Grid." *Agricultural Systems* 127: 53–60.

Zoltner, John. 2015. "The Future Is Here with Sensors and Geospatial Analytics in International Development." *ICTworks*. http://www.ictworks.org/2015/06/15/the-future-is earlier-here-with-sensors-and-geospatial-analytics-in-international-development.

GLOSSARY

2G, 3G, 4G. Second-, third-, and fourth-generation [developments in mobile wireless technology]. 2G mobile wireless has basic functionality: voice and short messaging service (SMS); 3G has advanced functionality: general packet radio service; and 4G has broadband functionality: long-term evolution (LTE).

Active infrastructure sharing. The shared use of electronic infrastructure such as network components (for example, access node switches), radio transmission equipment, and core network software systems. See *passive infrastructure sharing*.

Aerial photography and orthophoto mosaic. An image (once a photograph, now a digital image) of the ground taken from an airplane, helicopter, or radio-controlled aircraft at a given altitude. Aerial images are presented as an orthophoto mosaic that is an alternative to a map. These images are higher in resolution (decimeter) than satellite images, proving useful for those who want more details of the terrain such as crop conditions or land use.

Agricultural innovation system (AIS). A network of organizations, enterprises, and individuals focused on bringing new products, new processes, and new forms of organization into economic use, together with the institutions and policies that affect their behavior and performance.

Application. A software program or groups of programs enabling users to perform particular operations. They consist of systems software (operating systems for managing computer resources, for example) and programs such as those for data processing, word processing, and a multitude of functions that run on systems software. An IT application for managing dairy cooperatives, for example, relies on numerous kinds of applications running on the operating systems of any number of devices and the Internet. See http://www.webopedia.com/TERM/A/application.html.

Basis risk. In index-based insurance, the imperfect relationship between the policy holder's potential loss and the behavior of the index. One farmer's loss from drought may not perfectly match that of all others; some farmers will lose more and some less.

Biometric cards. Identification cards with a microchip or barcode that contains information on the physical characteristics of the holder. These cards can help prevent fraud and identity theft by providing a more accurate means of identification.

Broadband. Specifically, a signaling method that handles a relatively wide band (spectrum) of electromagnetic frequencies. More generally, the term refers to a telecommunications signal or device of greater bandwidth than another standard or usual signal or device (and the broader the band, the greater the capacity for traffic). The wider (or broader) the bandwidth of a channel, the greater its information-carrying capacity, given the same channel quality. (Based on http://en.wikipedia.org/wiki/Broadband#Internet_access, accessed July 2011.)

Chain traceability. Recording and transferring product or process data through a supply chain between various organizations and locations involved in the provenance of food. See *internal traceability*.

Cloud computing. A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud computing permits organizations without the resources to invest in extensive computing power to rent this service from a provider and access it remotely. (Based on http://en.wikipedia.org/wiki/Cloud_computing?oldid=0, accessed August 2011.)

Commercial supply chain. In agriculture, a supply chain in which a private agribusiness is sourcing agricultural produce from farmers or selling products to farmers in accordance with a profit-seeking business model. Often used interchangeably with supply chain and value chain.

Commodity futures exchange. A market in which multiple buyers and sellers trade commodity-linked contracts on the basis of rules and procedures set out by the exchange. Such exchanges typically act as a platform for trade in futures contracts (standardized contracts for future delivery of a commodity). (Based on a definition by the United Nations Conference on Trade and Development.)

Crowdsourcing. Shorthand for leveraging mass collaboration through ICTs by distributing tasks to or requesting information from a large group of people or community ("crowd") through an open call or message.

Data mediation. The process of using many data sets to produce a single, coherent set of information. Data mediation software organizes different types of data (such as hourly versus daily) and synthesizes different approaches to classification (for example, the use of a different classification vocabulary), helping to mediate differences between data sources—particularly those on the Internet.

Data mining. The extraction of stories or patterns from large amounts of data. Data mining can follow four major patterns: clustering (discovering groups), classification (forming a structure), regression (finding a function), and associations (finding relationships).

Digital divide. Differences in the capacity to access and use ICTs among individuals, men and women, households, geographic areas, socioeconomic groups, ethnic groups, and so forth. The capacity to access ICTs encompasses physical access as well as access to the resources and skills to participate effectively as a "digital citizen." (Based on the definition in http://en.wikipedia.org/wiki/Digital_divide, accessed July 2011.)

Digital orthophoto quads. Digital maps that combine the geometric information of a regular map with the detail of an aerial photograph.

Digital soil mapping. The creation and the population of a geographically referenced soil database generated at a given resolution through field and laboratory observation methods, coupled with environmental data through quantitative relationships. A variety of technologies—including satellite, remote sensors, and cameras—can be used to survey soil and collect data to create digital soil maps.

Digital terrain model. A digital representation of an area's terrain on a GIS that provides accurate position and elevation coordinates. Such models can be used to meticulously engineer projects such as roads, drainage, gravity-fed irrigation works, and detention reservoirs. At the field level, digital terrain models can monitor and improve areas affected by waterlogging or flooding.

e-government. A government's use of ICT to enhance public services.

e-Learning. is the use of electronic technologies to deliver, facilitate, and enhance both formal and informal learning and knowledge sharing at any time, in any place, and at any pace.

Elite capture. When better-off or politically connected farmers capture public programs.

Enterprise resource planning (ERP). Software integrates the many functions of an enterprise into a single system. It centrally stores many kinds of organizational data and manages data transmission and use between departments within the organization and external partners, such as suppliers. ERP is more of a methodology than a piece of software, although it does incorporate several software applications under a single, integrated interface.

e-readiness. The ability to use ICT to develop or improve one's economy or situation through proper preparation.

Farmer-led documentation (FLD). A process in which local communities take the lead role in the documentation process. The results are used by community members for learning within the community (internal learning); exchange between communities (horizontal sharing); and cooperation between communities, development agents, and policymakers (vertical sharing). See www.prolinnova.net/fld.php.

Feature phones. A modern low-end phone that is not a *smartphone*. Feature phones do not run a mobile operating system like smartphones but run on specialized software enabling them to access various media formats in addition to offering basic voice and SMS functionality. They substitute for multiple ICT devices that are also available as stand-alone appliances (digital camera, voice recorder, flashlight, radio, and MP3 player). Rural consumers prefer the combined devices because of their affordability. (Based on http://en.wikipedia.org/wiki/Feature_phone, accessed July 2011.)

Financial inclusion. The delivery of affordable financial services to disadvantaged and low-income segments of society. Research on financial exclusion and its direct correlation with poverty has made the availability of banking and payment services to the entire population without discrimination a prime objective of public policy. (Based on http://en.wikipedia.org/wiki/Financial_inclusion, accessed July 2011.)

Fixed-mobile convergence. The increasingly seamless connectivity between fixed and wireless telecommunications networks, devices, and applications. Also refers to any physical network that allows mobile phones to function smoothly with the fixed network infrastructure. FMC seeks to optimize transmission of all data to and among end users, no matter their locations or devices. (Based on the definition in http://searchmobilecomputing.techtarget.com/definition/fixed-mobile-convergence, accessed July 2011.)

Genetically modified (GM). A genetically modified organism (GMO) in which the genetic material has been transformed using the techniques of genetic engineering. Examples include cotton that has been genetically transformed to resist a particular herbicide. Many countries strictly control the production, use, export, and import of GM plants and animals.

Geographical information system (GIS). Geographic data collected through computer hardware and software to capture, store, update, and display all forms of geographically referenced information by matching coordinates and time to other variables. Data sets formed by a GIS constitute "layers" of information (for example, on topography, population size, or agricultural household income) that can be merged and analyzed to establish relationships and produce maps or charts that visualize geographical traits.

Georeference. To establish the position of something through its geographical coordinates.

Global positioning system (GPS). A satellite-based navigation system with three basic components: satellites that orbit the Earth, control and monitoring stations on the Earth, and the GPS receivers owned by users. GPS receivers pick up signals from the satellites, including precise orbital information (latitude, longitude, and ellipsoidal GPS altitude) of a given object or location, as well as the time.

ICT. Information and communication technology.

Index-based insurance. Insurance that substitutes individual loss assessments with an indicator that is easy to measure (such as weather) as a proxy for the loss. Weather indices have been used in insurance products protecting against drought and loss of inputs. Vegetation has been used in livestock insurance products as an indicator of livestock losses. See also weather-based index insurance and basis risk.

Infomediary. An infomediary works as a personal agent on behalf of consumers to help them take control over information gathered about them for use by marketers and advertisers. (Based on http://en.wikipedia.org/wiki/Infomediary, accessed September 2011.)

Internal traceability. Data recorded within an organization or geographic location to track a product or process. See *chain traceability.*

Laser scanning, or light detection and ranging (LiDAR). An active airborne sensor using a set of laser beams to measure distance from an aircraft to features on the ground. Airplanes and helicopters can be used for laser scanning. The data from laser scanning are three-dimensional at very high accuracy, and they also allow ground elevation under the tree canopy to be measured.

Market intelligence. Information relevant to the markets that a producer (or company) wishes to reach, which is gathered and analyzed specifically for making strategic decisions that will help to maximize profits in relation to market opportunities, market penetration, and market development. Market intelligence is necessary when entering a new market (foreign or domestic).

Mobile application. Software on a portable device (such as a mobile phone handset, personal digital assistant, or tablet computer) that enables a user to carry out one or more specific tasks that are not directly related to the operation of the device itself. Examples include the ability to access specific information (for instance, via a website), make payments and other transactions, play games, and send messages.

Nanotechnology. The ability to engineer new attributes by controlling features at or around the scale of a nanometer (one-billionth of a meter, or about 1/80,000 the width of a human hair).

Passive infrastructure sharing. The sharing of nonelectronic infrastructure, equipment, and services at mobile network base stations, including the site space, buildings, towers, masts, and antennas; power supply, back-up batteries, and generators; security; and maintenance.

Precision farming (precision agriculture). Farming based on observing and responding to variations within a field detected through ICTs, such as satellite imagery. Precision farming also makes use of GPS, GIS, and variable rate technology to match practices more closely to the needs of crops, soils, animals, or fisheries.

Primary wholesale market. A market large enough to dominate trade in some goods over a large area. (Based on http://www.merriam-webster.com/dictionary/primary%20market, accessed July 2011.)

Radio-frequency identification (RFID). Uses radio waves to transfer data between a reader and an electronic tag attached to a product, animal, or person for identification and tracking. The technology uses hardware (readers) and tags (also known as labels) as well as software. Most tags contain at least two parts: one is an integrated circuit for storing and processing information, and the other is an antenna for receiving and transmitting the signal. (Based on http://en.wikipedia.org/wiki/Radio-frequency_identification, accessed July 2011.)

Risk. Imperfect knowledge where the probabilities are known. Traditional risks to agriculture in developing countries include inclement weather, pests, disease, outbreaks, fire, theft, and conflict. Newer risks include commodity and input price volatility. Risks can be idiosyncratic—affecting only individual farms or firms—or covariate, affecting many farms and firms simultaneously.

Risk coping. Actions that help the victims of a risky event (such as a drought, flood, or pest epidemic) cope with the losses it causes. They include government assistance to farmers, debt restructuring, and remittances.

Risk mitigation. Actions that prevent events from occurring, limit their occurrence, or reduce the severity of the resulting losses (for example, pest and disease management strategies).

Risk transfer. Actions that transfer risk to a willing third party, at a cost. Financial transfer mechanisms trigger compensation or reduce losses generated by a given risk, and they can include insurance, reinsurance, and financial hedging tools.

Sanitary and phytosanitary (SPS) protection. Measures, including regulations and agreements, to protect: (1) human or animal health from risk arising from additives, contaminants, toxins, or disease organisms in food, drink, and feedstuffs; (2) human life from risks associated with diseases carried by plants or animals; (3) animal or plant life from pests, diseases, and disease-causing organisms; and (4) a country from other damage caused by the entry, establishment, or spread of pests. Such measures include national control of contaminants, pests, and diseases (vaccination programs, limits on pesticide residues in food) as well as international controls to prevent their inadvertent spread (for example, the rejection of insect-infested food shipments that pose a risk to domestic food production).

Satellite imagery. An image of Earth taken from satellites in orbit. Satellite imagery can be spatial (size of surface area); spectral (wavelength interval); temporal (amount of time); and radiometric (levels of brightness). Each type of images captures a variety of variables about a given area of varying size. The resolution (in meters) of these images depends on the satellite system used and its distance from Earth; weather can interfere mainly with satellite systems utilizing visible wavelengths of light.

Side-selling. A farmer sells produce to a buyer other than the agreed-on buyer. Farmers may fail to honor contracts with buyers for a number of reasons (buyers pay late, or prices in the local market are higher than the original price agreed on with the buyer, for example).

Smartcard. A pocket-sized (usually plastic) card with embedded integrated circuits containing volatile memory and microprocessor components. They include credit cards, identification cards, and the SIM cards used with mobile phones. As discussed in this sourcebook, one of their most influential roles has been to extend the use of mobile phones in financial transactions such as purchases of subsidized inputs, conditional cash transfers, agricultural credit, and agricultural information services. (Based on http://en.wikipedia.org/wiki/SmartCard#Cryptographic_smart_cards, accessed July 2011.)

Smartphone. A high-end mobile phone that offers more advanced computing ability and connectivity than a contemporary feature phone. A smartphone runs a complete mobile operating system and combines the functions of a personal digital assistant (PDA) and a mobile phone. Today's models typically serve as portable media players and camera phones with high-resolution touchscreen, global positioning system (GPS) navigation, Wi-Fi and mobile broadband access. (Based on http://en.wikipedia.org/wiki/Smartphone, accessed July 2011.)

SMS (short messaging service). A service to send text messages via mobile or fixed-line phones, usually limited to about 160 characters.

Soil carbon sequestration. Transferring carbon dioxide from the atmosphere into the soil through crop residues and other organic solids (like mulch), is one technique to restore carbon levels in soils.

Soil organic carbon. Carbon held within the soil as a result of the decay of once-living plants and animals. The amount of carbon within the soil is used as a measure of soil organic matter; soils with high levels of organic matter are better at holding water and contain more nutrients.

Spatial modeling (among other models). Closely related to spatial analysis or statistics, models are an attempt to simulate real-world conditions and explore systems using their geographic, geometric, or topological properties.

Spectrum rights. Rights to specific parts of the radio spectrum used for radio transmission technologies and applications. The radio spectrum is typically regulated by governments, and in some cases is sold or licensed to operators of private radio transmission systems (for example, cellular telephone operators or broadcast television stations). (Based on http://en.wikipedia.org/wiki/Radio_spectrum#Broadcasting, accessed July 2011.)

Subscriber identity module (SIM). An integrated circuit that securely stores the service-subscriber key used to identify a subscriber on mobile devices (such as mobile phones and computers). A SIM is held on a removable SIM card, which can be transferred between different mobile devices. (Based on http://en.wikipedia.org/wiki/Subscriber_Identity_Module, accessed October 2011.)

Supply chain. The set of buy-sell interactions as goods flow from raw materials through production to the final retailer where consumers can buy them. Often used interchangeably with commercial supply chain and value chain.

Supply-chain management (SCM systems). Software running on networked computers and handheld devices to perform some or all of the following functions: store information about suppliers; transmit an order to the supplier (in an agricultural supply chain, often the farmer); monitor production and quality; transfer payments; and track goods from the farm gate to the warehouse or retailer.

Technological neutrality. A leading regulatory policy principle for ensuring the affordability of ICTs, technology neutrality is the principle of refraining from specifying technology requirements within telecommunications licenses.

Telecenter. A public place where people can use digital technologies (computers, the Internet, even mobile phones) to gather information, create, learn, and communicate with others. Some centers are established specifically for people to learn these essential digital skills; others simply operate profit. But telecenters often help to support community, economic, educational, and social development—reducing isolation, bridging the digital divide, and creating economic opportunities. (Based on http://en.wikipedia.org/wiki/Telecenter, accessed July 2011).

Traceability (product tracing system). The information system necessary to provide the history of a product or a process from origin to point of final sale. Traceability is used in the food sector primarily for food safety, but agrifood and nonfood sectors such as forestry and textiles have instituted traceability requirements for product identification, differentiation, and historical monitoring. For food products, traceability systems involve the unique identification of products and the documentation of their transformation through the chain of custody to facilitate supply chain tracking, management, and detection of possible sources of failure in food safety or quality.

Uncertainty. Imperfect knowledge, where the probabilities are not known. Many losses expected from risks inherent in modern agrifood systems are related to uncertain events for which there are no known probabilities.

Universal access (UA). (also termed "public," "community," or "shared" access) occurs when everyone can access communications networks somewhere, at a public place. (Generally, the goal is to have at least one point of access per settlement over a certain population size.) As a policy objective, UA is used primarily in developing countries, which seek to expand geographic access to ICTs by the population at large, often for the very first time. UA obligations provide for a minimum level of coverage, especially of remote communities.

Universal service (US). A concept underpinning the definition of access to ICTs, US occurs when every individual or household can have service from communications networks, accessing services privately at home or increasingly through portable wireless devices. US focuses on upgrading and extending communication networks so that a minimum level of service is delivered, even in the least accessible areas. As a policy objective, US is used primarily in developed countries and generally pursued by imposing universal service obligations on network operators. For some services, a goal of US is too ambitious at present in a developing country, because the services must be affordable as well as available. Goals may be cast in terms of the proportion of the population that can afford private service.

Userability. The degree to which an ICT application is user friendly—a critical aspect of successful ICT implementation.

Value chain. The whole ecosystem of players involved in producing and marketing an article, from the retailer back to the producer. Often used interchangeably with commercial supply chain and supply chain.

Variable rate technology. Technology enabling farmers to vary the rate of an input applied to a crop. This technology uses a variable rate control system in combination with application equipment to supply inputs at the precise time and/or place where they are required. Components of the technology include a computer, software, differential GPS receiver, and controller. See *precision farming*.

Weather-based index insurance. Insurance that substitutes an indicator that is easy to measure for individual loss assessments (in this case, weather) as a proxy for the loss. Weather events or visible vegetation have served as typical indicators. This practice reduces the cost of assessing damage and problems of adverse selection, because the insured cannot influence the index or the loss assessment.

Web 2.0. Web 2.0 sites (unlike websites where users passively view content) incorporate applications that facilitate participatory information sharing, interoperability, user-centered design, and collaboration through the Internet. Examples include social networking sites, blogs, wikis, video sharing sites, and hosted services. (Based on http://en.wikipedia.org/wiki/Web_2.0, accessed September 2011.)

WiFi. Wireless local area network that allows various devices to connect to the Internet remotely.

Wireless sensor network. A group of small sensing devices, or nodes, that capture data in a given location and send it to a base station in the network, which transmits the data to a central computer that performs analyses and extracts meaningful information.

Information and communication technology (ICT) has always mattered in agriculture. Ever since people have grown crops, raised livestock, and caught fish, they have sought information from one another. Today, ICT represents a tremendous opportunity for rural populations to improve productivity, to enhance food and nutrition security, to access markets, and to find employment opportunities in a revitalized sector. ICT has unleashed incredible potential to improve agriculture, and it has found a foothold even in poor smallholder farms.

ICT in Agriculture, Updated Edition is the revised version of the popular ICT in Agriculture e-Sourcebook, first launched in 2011 and designed to support practitioners, decision makers, and development partners who work at the intersection of ICT and agriculture. Our hope is that this updated Sourcebook will be a practical guide to understanding current trends, implementing appropriate interventions, and evaluating the impact of ICT interventions in agricultural programs.

KWPF
KOREA-WORLD BANK PARTNERSHIP FACILITY

WORLD BANK GROUP

infoDev

ISBN 978-1-4648-1002-2

90000

9 781464 810022

SKU 211002