A researcher uses a handheld GPS device in Colombia. The instrument will help researchers collect and aggregate data on the planting decisions of hundreds of the country's fruit farmers. (Neil Palmer/CIAT)

14

Big Data in Agriculture and Nutrition

Ruthie Musker*
Global Open Data for Agriculture and Nutrition (GODAN), Wallingford, UK

* ruthie.musker@godan.info

Why Do Agriculture and Nutrition Need Big Data?

The earth provides humans with enough resources to feed our growing population, yet 815 million people live with chronic hunger. Although human health has generally improved, the current food system faces many challenges, including undernutrition, micronutrient malnutrition, and rising rates of obesity (FAO, 2018).

The amount of data collected on global food systems is immense, and the Sustainable Development Goals (SDGs) of the United Nations (UN) (especially SDGs 2, 3 and 17) encourage the sharing of information and data on agriculture and nutrition. Despite this call to action, many stakeholders, from farmers to governments, lack actionable data-driven insights and a clear understanding about how data translate into action. There are still many knowledge gaps on linkages among agriculture, nutrition, and the food system, especially complex systemic issues throughout the value chain. The power of data to begin closing these gaps remains largely untapped.

The food system community sees a huge potential for big data in agriculture to lift farmers out of poverty (Patel, 2013), and ensure that parents can feed their children nutritious, diverse foods (Lung'aho, 2018). In the USA, venture capitalists spent US$3 billion on 'agtech' (digital technology in agriculture) in 2016, with 46% of investors focusing on big data and analytics (Walker *et al.*, 2016). Large data initiatives such as the CGIAR's Big Data in Agriculture Platform have made thousands of datasets and publications available (Pineda, 2018). In order to establish a global data ecosystem that yields powerful insights and recommendations on the ways in which agriculture can improve nutrition, the community must ensure that the benefits of big data are for the betterment of all and not only for the few.

What Is Big Data?

Is big data a trend hyped by the media, or does it indeed have the power to 'disrupt' agriculture systems for the benefit of nutrition? Like many terms within agriculture, nutrition, and sustainable development such as 'food security' (Gibson, 2012), and 'food system' (e.g. Edgar and Brown, 2013), 'big data' lacks a universally agreed definition (Bhadani and Jothimani, 2016).

The overarching characteristics of big data that apply to most disciplines are the 3Vs: Volume, Velocity, and Variety, with a fourth V, Veracity, also applicable to agriculture and nutrition.

- **Volume:** how much data are collected. Volume depends on the amount over time, which informs the next component (Bhadani and Jothimani, 2016).
- **Velocity:** how fast data are collected. In agriculture and nutrition, an oft-mentioned benefit of big data is the opportunity for near-real-time analysis and decision-making. For example, early warning systems provide real-time data on agricultural production, weather patterns, nutritional status, and other factors and send alerts to policy-makers on emerging humanitarian crises.
- **Variety:** what types of data are collected. Variety of data is one component that makes big data especially applicable to agriculture and nutrition. With the onset of digital data collection, the internet, and smartphones, big data has changed what data 'look like'. Instead of numbers on crop yields or stunting rates in a spreadsheet, data also include maps and GPS coordinates, photos (of eating habits, for example), texts (nutrition messaging), relationships (mapping of agriculture–nutrition stakeholders, for instance) and many more (Sonka, 2014).
- **Veracity:** how reliable is the data source. Good data quality is essential for optimal decision making, especially when one decision can impact the nutritional status or livelihoods of a large segment of a population (Gandomi and Haider, 2015).

It is important to consider that big data fulfills a specific role within the larger data ecosystem. The data ecosystem includes all sizes and types of datasets. Data may not become big data unless they are analyzed at a certain scale. Importance and impact of the dataset may not be correlated with the dataset size.

What Does Big Data Look Like in Agriculture and Nutrition?

Data are collected from a variety of sources and in many ways, which is why applications of big data can be applied in so many ways across the food system. As mentioned earlier, in order for data to be 'big', there must be a large amount, collected quickly, that takes a wide variety of forms. There are many datasets within agriculture and nutrition that fulfill these criteria, as listed in Table 14.1.

Table 14.1. Types of data in agriculture and nutrition (scale will determine whether or not they are big data).

Data type	Sub-type	Definition	Sources
Remote sensing		The collection of information from a distance	NOAA, 2018
	In situ (subsurface) sensors	Small scale, stationary, attached to the earth, such as weather stations or water quality sensors *Example*: Satellites use weather stations to validate and enrich data sources to provide accurate, real-time information to farmers	Kotamäki *et al*., 2009
	Aerial, non-satellite	Medium scale, sensors on aircraft, such as Unmanned Aerial Vehicles (UAV), drones *Example:* A company called Agribotix is developing drones with sophisticated multispectral sensors and programs that detect pests and diseases and deliver quick targeted solutions automatically	Clause *et al*., 2018 King, 2017
	Satellite	Large scale, sensors on satellites *Example:* The company Planet images the Earth every day, collecting 1.4 million 29MP images per day, covering more than 300 million km^2. Over 6 terabytes per day are sent back to earth *Example:* The European Space Agency's Copernicus satellites collect atmospheric and climate data open for public use. The data inform programs like APOLLO, which provides advisory services for small farms	Clause *et al*., 2018 Planet, 2017 Copernicus, 2018
Farm equipment and robotics		Farm equipment equipped with GPS, guidance systems, crop-specific sensors, for planning, monitoring, analysis and planning. Also called, 'smart farming' and 'precision agriculture' *Example:* Powerful algorithms allow robots to use the RGB spectrum to pick strawberries precisely at peak nutritional value. Big data is used to power the algorithm, and every strawberry picked becomes a data point in a big dataset *Example:* A small autonomous robot called Bonirob can analyze soil samples to map pH and phosphorus concentrations in real time, helping to improve soil health	Killpack, 2011 Wolfert *et al*., 2017 King, 2017
Mobile phones through social media and crowd-sourcing		Mobile phones have allowed for collection and dissemination of information on a very large scale *Example:* Mobile phones allow smallholder farmers to share to information and receive alerts and recommendations around planting and selling	USAID, 2013 Noronha *et al*., 2011

Continued

Table 14.1. Continued.

Data type	Sub-type	Definition	Sources
		Example: Fitness and diet tracking applications allow companies to observe eating habits of a population on a wide scale. Apps like Twitter and Facebook (which also owns Instagram and WhatsApp) can view their users' messages, photographs, and behavioral trends, such as eating habits	
Omics		The basis of human health, nutrition and disease knowledge, and is also applied to agricultural science through pest and disease resistance, and nutritional value of crops *Example:* Personalized medicine and personalized diets can be developed from omics big data generated by universities, industry, combined with an individual's personal health data, either from their hospital's health records or the health apps attached to mobile phones (Apple HealthKit, Samsung S-Health) and wearables (Fitbit, Apple Watch)	Alyass *et al.*, 2015 Van Emon, 2016 Kraft, 2017
Research		Any type of research data collected by industry, government, or academia, including market research collected by private-sector companies on trade, marketing, wholesaling, retailing, and online sales	

While big data can be sourced from industry, academia, and government, it can also be generated by the users of farm equipment, mobile phones, and social media. When people use an app, the information they input and their behavior while using the app then becomes big data for others to interpret and use. As the number of mobile phones and smartphones increases, the data that is generated also increases. Although it is difficult to find examples of user-generated big data for nutrition, apps such MyFitnessPal and other diet trackers may provide examples in the future.

Two of the biggest challenges for farmers are risks from external stresses and a lack of a safety net. Early-warning systems and insurance help farmers overcome these risks, especially in the age of climate change. The use of big data is allowing for better early warning systems and insurance schemes than ever before, using a combination of data types. Initiatives such as the Famine Early Warning System Network (FEWS NET) can drastically improve evidence-based analysis for decision-making in the most vulnerable places. FEWS NET was established in 1985 and, using big data from satellites and research, it can publish reports and maps of food insecurity projections, as well as crisis alerts and specific data on weather, markets, and nutrition, allowing governments to help citizens in a timely way (FEWS NET, 2018). The impact of FEWS NET could be better assessed if research were collected on how governments used the early warning system and how lives of the affected people were improved.

Big data can also help with insurance and access to credit with a combination of big data types. India-based company Satsure analyzes satellite data, market data, and weather data using machine learning and big data analytics to ensure that farmers in India who have suffered crop losses due to climatic shocks receive compensation quickly. Satsure is a relatively new company and evaluations of its program are forthcoming (e-Agriculture, 2017).

From Data to Decision

All types of big data must go through a series of steps, in combination with other types of data,

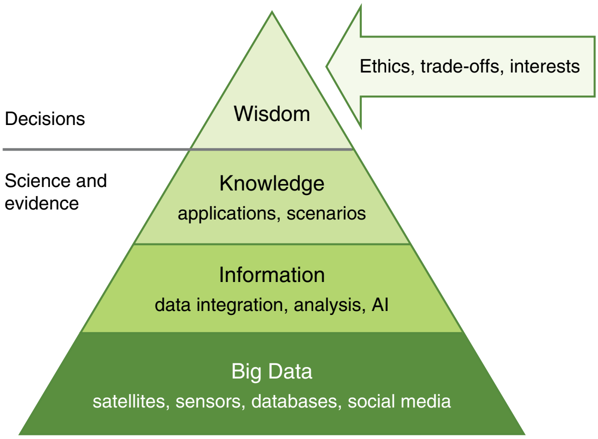

Fig. 14.1. DIKW hierarchy, from big data to decision-making for societal challenges (adapted from Lokers *et al*., 2016).

with a variety of stakeholders in order to make a data-driven decision. Figure 14.1 shows this potential pathway.

Big data from the sources mentioned above will contribute to the development of an evidence base. Data must be analyzed by companies or individuals who have developed tools to do so, and after analysis several options will be presented for certain scenarios and applications. Big data may also be used to create software to predict and quickly determine where a problem is happening and tell the user the best way to overcome it. In order for that software to be written, large amounts of data are needed to support the recommendations (Lokers *et al*., 2016). It is only after this step that decision makers have the knowledge to make a decision. Their own ethical position, interests, and interpretation of potential trade-offs will in turn influence the final result.

Challenges in Big Data in Agriculture and Nutrition

Several challenges must be overcome in order for big data in agriculture and nutrition to help stakeholders make optimal decisions.

Technological challenges

By definition, 'big data' is very large and complex data, often requiring high-end, extensive (and costly) technology for management and analysis. Food systems research is increasingly interdisciplinary, which makes data management a more complex challenge than other domains. Each discipline will have different targeted objectives, data formats, schemas, vocabularies, standards, and granularities (Lokers, 2015).

As more data are gathered on agriculture and nutrition, the 3Vs of big data – volume, velocity, and variety – will increase exponentially. In order to make best use of the data, large investments will be needed to store and preserve the data on platforms and in databases. Time, effort, and money to spend on data management technology are currently minimal across relevant stakeholders (Shekhar *et al*., 2017).

Data are collected in different ways, in different formats, using different technologies and in different languages. For big data to be most effective, different data should be able to be layered on top of each other, with each layer helping to further inform a solution or decision. This layering is referred to as 'data integration' with the data needing to be 'interoperable' for data integration to occur. Data integration can be fairly easy if the data sources are interoperable, as XML, APIs or text files (Kadadi *et al*., 2014), but if they are not, or the semantics and vocabularies do not match, data integration may be difficult, expensive, or impossible.

Institutional challenges

Arguably, high data quality supersedes any other analysis or integration issue, as the costs of bad-quality data may be greater than having no data at all (Cai and Zhu, 2015). High data quality is also essential for building trust in data sharing (Allemang and Teegarden, 2016). Quality assurance standards are not common in agriculture and nutrition data collection or management. However, it is also difficult to know 'how good is good enough?' Progress is being made in this area (Grassini *et al.*, 2015; Lu *et al.*, 2015).

For almost all stakeholders, institutional data management is an afterthought. It is usually not integrated into the research design or collected with the intent to share or reuse it. Money is not allocated for quality assurance, curation, and sustainability. Depending on the research project, the volume, velocity, and variety of data may make retroactive data management difficult (Smith *et al.*, 2017; Adrian *et al.*, 2018; Roett, 2018).

Cultural challenges

The standard operating procedure of business, science, and management is closed data, meaning data that are not open or shared (ODI, 2015). If big data is to be used optimally, organizations need to share or open their data. However, this process may require them to change their business models, the people they hire, their business relationships, and their institutional culture. Such a process is slow and potentially threatening to risk-averse organizations, or those that do not have the financial or human capacity to change. Researchers in universities are especially averse to opening and sharing data, for fear of others stealing their results. However, they are open to reusing data that others have published (Digital Science, 2017). Other cultural considerations include bureaucracy and other social structures that impede data sharing, norms and structures that can be highly variable across countries or regions.

Ethical challenges

Data ownership rights are usually absent in legal frameworks over the handling of agriculture and nutrition data. More often, data are owned by the person or organization that collects them (or the one that funds their collection) due to a proprietary interest in the data being collected, instead of the person that the data is about. This can lead to privacy and security issues, along with the emergence of a digital divide, meaning that big data is helping powerful entities, instead of improving livelihoods of the disadvantaged. Most smallholder farmers are not able to understand, interpret, and use the analysis of data without intermediation. With the increase of smartphones, GPS on tractors, wearable technology or devices, and personally identifiable information, these ethical challenges are crucial to overcome (de Beer, 2016; Kshetri, 2014).

An example of this ethical challenge can be seen within the context of the operating practices of firms that manufacture tractors and heavy agricultural tools. One such manufacturer, John Deere, has made tremendous advances in precision farming using crowd-sourcing and remote sensing in the USA. The company tracks its agricultural machinery on each farm and aggregates it to improve predictions and provide recommendations, usually to promote its own products. However, since John Deere also owns the data that its machines generate, farmers cannot see the data being sent to John Deere unless they buy it back. Some farmers believe that they should own the data and be compensated if John Deere uses it to make business decisions (Woodard *et al.*, 2017). In a survey performed by Farm Industry News, farmers expressed a desire to be in control of their data and were concerned about how their data may be used (Farm Industry News, 2016).

Lessons Learned and Solutions Towards Putting Big Data to Use in Agriculture and Nutrition

Recent research and current initiatives are leveraging the potential of big data to help solve large problems in agriculture and nutrition (Kshetri, 2014), though not enough time has yet passed to see genuine, sustained benefits. However, big data's momentum is forcing a wide range of stakeholders to learn from one another to develop innovative and novel solutions to the challenges listed above.

Internet of Things (IoT)

The Internet of Things (IoT) has a huge potential to connect agriculture and nutrition data, providing insights on how nutrition can be retained along the food value chain. The IoT aims to interconnect objects such as mobile phones, tractors, *in situ* sensors, and wearables using wireless sensors, radio frequency identification (RFID), and other web-based capabilities, and tackles the data integration challenge. From the agricultural production angle, the IoT would provide the tools to better monitor agricultural production by providing a smarter understanding of farming conditions, rainfall, pest and disease threats, and best management practices. It lays the groundwork for high-tech, remote-controlled farm logistics and processing, such as robots for weeding and precise fertilizer application. The IoT would then link production to logistics by remotely monitoring ambient conditions during transportation, positively impacting food quality and traceability. Subsequently, the IoT could combine the results of the IoT chain with personalized health through wearables, omics data, mobile phone apps, and documented nutritional data from healthcare providers, documenting the link between production and nutritional status (Sundmaeker *et al.*, 2016). Data must be interoperable for a successful IoT to develop.

Open data

Data should drive all important decisions in agriculture and nutrition, big or small (see Fig. 14.1). Open data is data that anyone can access, use, or share (ODI, 2018) and is potentially the most impactful way that big data can make a difference in agriculture and nutrition. In addition to fast and effective decision making, open data can drive innovation that everyone can benefit from, and can promote organizational and sector change through transparency (Carolan *et al.*, 2015). Agriculture and nutrition are highly interdisciplinary, and open data will allow stakeholders to more easily access and use data from previously inaccessible disciplines. Research and support behind open data is strong (Allemang and Teegarden, 2016), but there are knowledge gaps, in terms of clear examples of how opening data can explicitly overcome development challenges.

In order to maintain the high quality of open data, standards are needed. These provide guidelines on how to collect, manage, and integrate data and include common semantics and ontologies (Pesce *et al.*, 2018). One such standard is the FAIR Principles (**F**indable, **A**ccessible, **I**nteroperable, and **R**eusable), which are becoming more well known and accepted among researchers, governments, and other stakeholders. 'Findable' means that data can be found and curated; 'accessible' means that the data are usually in machine-readable code, or easily processed by a computer such as through XML or CSV; 'interoperability' allows data to be manipulated and aggregated with data from elsewhere to produce results that are of practical use; and 'reusable' means that the dataset should be openly licensed (Wilkinson *et al.*, 2016). Licensing provides guidelines on how the data can be reused. Most open datasets use the Creative Commons licensing system (Creative Commons, 2018).

The community using the FAIR Principles is growing and includes donors, universities, and governments, including the European Commission (EC) (DTL, 2016). The EC has developed 'Guidelines on FAIR Data Management in Horizon 2020' (European Commission, 2016) which mandates that all data from its Horizon 2020 projects, including those on food security, are open by default and adhere to the FAIR principles.

Collaborative platforms for big and open data

Organizations have learned that the speed of innovation depends on collaboration and mutual support. Several new initiatives are helping the food-system community collaborate and convene around the big data challenges and solutions.

The Big Data in Agriculture Platform is an initiative launched by CGIAR in 2017. The platform was created to overcome the challenge of big data management and the transformation of information into action. Its vision is to: organize existing data; improve data management, data generation, and access across the 15 CGIAR centers; convene members of CGIAR and its partners to use big data to solve agriculture and nutrition issues; and inspire others to do the same. The platform aims to achieve this vision by 2022 (CGIAR, 2018). To date, 2000 datasets and 50,000 publications have been made available (Pineda, 2018).

The Global Open Data for Agriculture and Nutrition (GODAN) initiative is a global network of over 850 (as of November 2018) partner organizations from all sectors, who advocate for open data and work together to overcome challenges, especially as they relate to food security. GODAN encourages all partners to open up key datasets, and to create policies for sustainable data sharing. The GODAN Partner Network includes organizations from all stages of the food system who have the opportunity to collaborate and see how their data can help others in the community. A primary goal of GODAN is trust-building and responsible open data management among partners (GODAN, 2018).

What Can Stakeholders Do to Make Big Data Work for Agriculture and Nutrition?

All stakeholders

Big data, when analyzed and layered together with other datasets within the data ecosystem, may help stakeholders in agriculture and nutrition to make better decisions across the entire food system. Although there are actions specific stakeholders can take towards making big data work for agriculture and nutrition, some actions are universal.

Collaboration

As in the example above of IoT, stakeholders from all sectors need to collaborate, share data, and co-strategize towards a common goal. For big data to have sustainable benefits for everyone, the key is cooperation and collaboration.

Responsible data use

A plethora of research exists on why data ownership and responsibility are important in big and open data in agriculture and nutrition (Kshetri, 2014; Bronson and Knezevic, 2016; Carbonell, 2016). If data are to be published and used responsibly to prevent power imbalances, empower vulnerable communities, and promote sustainable agriculture and nutrition (Ferris and Rahman, 2016), policies around clear privacy, security, and ownership principles must be drafted and consistently updated (de Beer, 2016); data subjects must be educated on how the data about them will be used and how they are compensated; and the rights of vulnerable people, especially smallholder farmers, must be protected.

Although the development community might broadly support these principles, there are as of yet few examples of its adherence in on-the-ground applications. Resources such as 'The Data Ethics Canvas' (ODI, 2017) can help ensure that responsible data use principles are followed.

Policies

Big data management, technology, and decision-making processes are relatively new, and policies are the best way to ensure that different sectors, regions, and disciplines have a joint understanding of the issues, cooperate on potential solutions, and produce common standards. Resources such as 'Writing a Good (Open) Data Policy' (ODI, 2016) exist for policy support across all sectors. Sector-specific policy suggestions and progress are given below.

Governments

Most governments across the world have ministries of agriculture, food, and health that collect and organize a tremendous amount of data. Governments are often the stewards of the data that they collect (Smith and Jellema, 2016), can own the data, and host it. Much of the data that exist across the world collected by governments may not be considered big data, especially within developing countries. However, governments have a responsibility to interpret big data and act upon it for the benefit of their citizens.

Governments can facilitate the information flows between their ministries and ensure high quality data by continuously cleaning, curating, and updating government data, as well as publishing open data on the web when appropriate. They can: (i) reinforce the national technical infrastructure so that the open data can be accessed easily and reliably at all times by other stakeholders; (ii) build the capacity among stakeholders to use big (and open) data sources; (iii) financially support stakeholders that want to build information services for the agricultural and nutrition sectors based on open data sources; (iv) encourage business development for sustaining the information services being developed; and (v) stimulate other stakeholders (e.g. private sector, international organizations, NGOs, researchers) to publish their own data sources (GODAN, 2018).

Several governments are making progress on big data and open data. In 2017, the Ministry of Agriculture in Kenya, for example, signed the Nairobi Declaration along with nine other African ministers, a public commitment to work jointly on open data in agriculture and nutrition and data-driven decision-making (GODAN, 2017).

Research organizations and universities

Research organizations and universities generate big data, but historically researchers are driven to publish articles in peer-reviewed journals, instead of releasing high-quality datasets. This mentality is beginning to shift, with researchers increasingly expressing interest in publishing datasets, as long as they are attributed and receive a citation. The main drive behind this interest is universities incentivizing dataset publication as they start to consider its contribution to publication counts and increased likelihood of new donor funding (Digital Science, 2017).

Data ownership is a point of contention between universities and researchers. Universities believe that if research is conducted on their campus (regardless of funding support), they own the data. Researchers disagree and believe that they themselves own the data. Universities must clearly define data ownership by collaborating with faculty and researchers on ownership policies. They should also create the infrastructure, support, and resources for data best practice and management (Adrian *et al*., 2018).

Donors

Donors spend billions of dollars on agriculture and nutrition research per year, funneled through universities, other governments, NGOs, and industry, all of which produce large amounts of data. Few donors fund the curation and maintenance of the high-quality data that results from their initial investments (Smith *et al*., 2017). Universities and research institutions, in turn, follow the data policies of their project funders. However, donors are only recently recognizing that although they may have data management policies (such as open data or open access), grantees often do not have the resources or knowledge to comply fully with them.

Donors can best support their grantees through a combination of compliance and incentives. They can regularly monitor compliance and articulate clear expectations regarding budget allocations to ensure good data management (Smith *et al*., 2017).

Industry

Each step within the food system (inputs, production, harvest, transport, storage, processing, retailing, consumption, and waste) has industry data collectors and users.

Publishers of academic research, such as Elsevier and Springer Nature (Springer Nature, 2016), are also industry stakeholders as they provide the primary throughput of scientific information that would lead to knowledge and decisions. Publishers can decide their own open data and open access policies, and pricing scales, to which researchers must adhere.

Business models for big data are well defined, but less so for big and open data. Companies such as Syngenta have found ways to publish open data for transparency and accountability. Syngenta's Good Growth Plan outlines six commitments for agricultural sustainability and has published the data for most of these components, including crop productivity, smallholder outreach, soil maintenance practices, biodiversity practices, and workplace safety. Syngenta found more value in making data open in promoting its social responsibility agenda than it would have in keeping the data closed (Allemang and Teegarden, 2016). Agribusiness may have the biggest challenge to address around data ethics policy, to ensure that farmers are not exploited for their data (Carbonell, 2016). More research and efforts are needed on development of sustainable business models for big open data in food systems.

In addition to business models, industry can adopt a view of data as a raw material to ascertain value for the company. The data value chain perspective can help companies use and reuse data to maximize analytics and tools for solving development problems and scaling up solutions (Dunhill, 2014). The research behind data value chains from IBM is solid, but is not yet applied in situational contexts.

Looking Ahead

As the international community works to fulfill the SDGs, big data will drive many of the efforts tied to linking agriculture and nutrition and reshaping the global food system. The collection of high-quality data is not sufficient. This vast well of information must translate into knowledge that is easily accessible by non-technical audiences, including policy-makers and civil society. By carefully building a system for open and big data, one that includes clear definitions, rules over ownership and use, and transparency and accountability, we can ensure that the benefits of big data are passed on to the most vulnerable segments of society.

Acknowledgements

I would like to thank Andre Laperriere, Martin Parr, Valeria Pesce, Ben Schaap, and Michael Brobbey for their valuable contributions to this chapter.

References

Adrian, A.M., Emison, B., Musker, R. and Hopkinson, J. (2018) *Open Access and Open Data at PUSH Universities*. Presidents United to Solve Hunger (PUSH) and Global Open Data for Agriculture and Nutrition (GODAN). Available at: http://wp.auburn.edu/push/wp-content/uploads/2018/06/GODAN-report.pdf (accessed 21 September 2018).

Allemang, D. and Teegarden, B. (2016) *A Global Data Ecosystem for Agriculture and Food*. Global Open Data for Agriculture and Nutrition (GODAN), Wallingford, UK.

Alyass, A., Turcotte, M. and Meyre, D. (2015) From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Medical Genomics* 8(1), 33.

Bhadani, A. and Jothimani, D. (2016) Big data: challenges, opportunities and realities. In: Singh, M.K. and Kumar, D.G. (eds) *Effective Big Data Management and Opportunities for Implementation*. IGI Global, Hershey, Pennsylvania.

Bronson, K. and Knezevic, I. (2016) Big Data in food and agriculture. *Big Data and Society* 3(1). doi: 10/1177/2053951716648174.

Cai, L. and Zhu, Y. (2015) The challenges of data quality and data quality assessment in the big data era. *Data Science Journal* 14, 2.

Carbonell, I. (2016) The ethics of big data in big agriculture. *Internet Policy Review*, 5(1). Available at https://ssrn.com/abstract=2772247 (accessed 19 December 2018).

Carolan, L., Smith, F., Protonotarios, V., Schaap, B., Broad, E., Hardinges, J. and Gerry, W. (2015). *How can we improve agriculture, food and nutrition with open data?* Open Data Institute (ODI), London.

CGIAR (2018) *Platform for Big Data in Agriculture*. CGIAR, Montpellier, France. Available at: http://bigdata.cgiar.org/ (accessed 19 December 2018).

Clause, V., Taverner, D. and Hayward, T. (2018) *UK Space Agency International Partnerships Programme: Space Solutions for Agriculture in Developing Countries*. Caribou Digital, Gainsborough, UK, and UK Space Agency, Swindon, UK.

Copernicus (2018) *Project Database: APOLLO*. Available at: http://copernicus.eu/projects/apollo (accessed 3 April 2018).

Creative Commons (2018) *About the Licenses*. Available at: https://creativecommons.org/licenses/ (accessed 8 March 2018).

de Beer, J. (2016) *Ownership of Open Data: Governance Options for Agriculture and Nutrition*. Global Open Data for Agriculture and Nutrition (GODAN), Wallingford, UK.

Digital Science (2017) *The State of Open Data 2017 Report. A selection of analyses and articles about open data, curated by Figshare*. Digital Science, London. Available at: https://www.digital-science.com/resources/portfolio-reports/state-open-data-2017 (accessed 28 September 2018).

DTL (2016) *GO FAIR Initiative*. Available at: https://www.dtls.nl/fair-data/go-fair/ (accessed 8 March 2018).
Dunhill, E. (2014) *Understanding the Data Value Chain*. IBM Big Data and Analytics Hub, November 10. Available at: http://www.ibmbigdatahub.com/blog/understanding-data-value-chain (accessed 9 March, 2018).
e-Agriculture (2017) *SatSure uses big data analytics and satellite imagery to help farmers in the agricultural insurance sector*. Available at: http://www.e-agriculture.org/news/satsure-uses-big-data-analytics-and-satellite-imagery-help-farmers-agricultural-insurance (accessed 1 April 2018).
Edgar, C. and Brown, L. (2013) *A Review of Definitions of Community Food Systems*. UW Cooperative Extension, University of Wisconsin–Extension, Madison, Wisconsin.
European Commission (2016) *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 Version 3.0*. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 8 March 2016).
FAO (2018) *What we do*. Food and Agriculture Organization, Rome. Available at: http://www.fao.org/about/what-we-do/en/ (accessed 8 March 2018).
Farm Industry News (2016) *Data concerns remain top of mind for farmers*. Available at: http://www.farmindustrynews.com/farm-equipment/data-concerns-remain-top-mind-farmers (accessed 8 March 2017).
Ferris, L. and Rahman, Z. (2016) *Responsible Data in Agriculture* [version 1; not peer reviewed]. *F1000 Research 2017* 6, 1306.
FEWS NET (2018) *Famine Early Warning System*. Available at: http://www.fews.net/about-us (accessed 1 April 2018).
Gandomi, A. and Haider, M. (2015) Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management* 35 (2), 137–144.
Gibson, M. (2012) Food security – a commentary: what is it and why is it so complicated? *Foods* 1(1), 18–27.
GODAN (2017) *GODAN supports historic Nairobi Declaration*. November 4. Available at http://www.godan.info/news/godan-supports-historic-nairobi-declaration (accessed 8 March 2018).
GODAN (2018) *Government Open Up Guide for Agriculture*. Available at: https://openupguideforag.info/ (accessed 21 September 2018).
Grassini, P., van Bussel, L.G., Van Wart, J., Wolf, J., Claessens, L. *et al*. (2015) How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crops Research* 177, 49–63.
Kadadi, A., Agrawal, R., Nyamful, C. and Atiq, R. (2014) Challenges of data integration and interoperability in big data. In: *IEEE International Conference on Big Data*, Washington, DC, 2014. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 38-40.
Killpack, C. (2011) Big data, bigger opportunity. *Geospatial World* 1(9), 18–26.
King, A. (2017) Technology: the future of agriculture. *Nature* 544, S21–S23. doi: 10.1038/544S21a.
Kotamäki, N., Thessler, S., Koskiaho, J., Hannukkala, A.O., Huitu, H. *et al*. (2009) Wireless in-situ sensor network for agriculture and water monitoring on a river basin scale in southern Finland: Evaluation from a data user's perspective. *Sensors* 9(4), 2862–2883.
Kraft, M. (2017) *Quantified self to quantified health: how tech helps doctors fill gaps in patient records*. 9 March 2017. Available at: https://www.wired.co.uk/article/hospital-prescribiing-tech (accessed 9 March 2018).
Kshetri, N. (2014) The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data and Society,* 1(2).
Lokers, R. (2015) *Big Data challenges and solutions in agricultural and environmental research.* Webinar@AIMS, 17 December. Available at: http://aims.fao.org/capacity-development/webinars/webinaraims-big-data-challenges-and-solutions-agricultural-and (accessed 8 March 2018).
Lokers, R., Knapen, R., Janssen, S., van Randen, Y. and Jansen, J. (2016) Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling and Software* 84, 494–504.
Lu, Y., Nakicenovic, N., Visbeck, M. and Stevance, A.S. (2015) Five priorities for the UN sustainable development goals. *Nature* 520(7548), 432–433.
Lung'aho, M. (2018) *I want to end malnutrition in Africa. That's why I'm turning to big data and algorithms*. CGIAR Blog. March 5. Available at: http://bigdata.cgiar.org/i-want-to-end-malnutrition-in-africa-thats-why-im-turning-to-big-data-and-algorithms/ (accessed 8 March 2018).
NOAA (2018) *Remote Sensing*. National Ocean and Atmospheric Administration, Silver Spring, Maryland. Available at: https://oceanservice.noaa.gov/geodesy/remote-sensing/ (accessed 8 March 2018).
Noronha, J., Hysen, E., Zhang, H. and Gajos, K.Z. (2011) Platemate: crowdsourcing nutritional analysis from food photographs. In: *Proceedings of the 24th Annual ACM Symposium on User Interface*

Software and Technology. October 16–19, 2011, Santa Barbara, California. Association for Computing Machinery (ACM), New York, pp. 1–12.

ODI (2015) *The Data Spectrum*. Open Data Institute, London. Available at: https://theodi.org/about-the-odi/the-data-spectrum/ (accessed 8 March 2018).

ODI (2016) *How to Write a Good Open Data Policy*. Open Data Institute, London. Available at: https://theodi.org/article/how-to-write-a-good-open-data-policy/ (accessed 8 March 2018).

ODI (2017) *The Data Ethics Canvas*. Open Data Institute, London. Available at: https://theodi.org/article/data-ethics-canvas/ (accessed 8 March 2018).

ODI (2018) *The Open Data Institute*. Open Data Institute, London. Available at: https://theodi.org (accessed 8 March 2018).

Patel, P. (2013) *Feeding the world with Big Data*. IEEE Spectrum (Institute of Electrical and Electronic Engineers). Available at: http://spectrum.ieee.org/computing/networks/feeding-the-world-with-big-data (accessed 8 March 2018).

Pesce, V., Kayumbi, G.W. and Tennison, J. (2018) *Agri-food Data Standards: a Gap Exploration Report* [version 1; not peer reviewed]. *F1000Research 2018*, 7, 176. Available at: https://f1000research.com/documents/7-176 (accessed 28 September 2018).

Pineda, S. (2018) *Big data revolutionising agriculture to combat world hunger*. CGIAR Blog. Available at: http://bigdata.cgiar.org/big-data-revolutionising-agriculture-to-combat-world-hunger/ (accessed 9 March 2018).

Planet (2017) *Mission 1 Complete!* Available at: https://www.planet.com/pulse/mission-1/ (accessed 8 March 2018).

Roett, K. (2018) *For a food secure future, we first need to look at how we manage data*. CGIAR Blog. February 19. Available at: http://bigdata.cgiar.org/for-a-food-secure-future-we-first-need-to-look-at-how-we-manage-data/ (accessed 8 March, 2018).

Shekhar, S., Schnable, P., LeBauer, D., Baylis, K. and VanderWaal, K. (2017) *Agriculture Big Data (AgBD) Challenges and Opportunities From Farm To Table: A Midwest Big Data Hub Community Whitepaper*. Available at: https://www.semanticscholar.org/.../Agriculture-Big-Data...AgBD...Shekhar-Schnable/c8... (accessed 28 September 2018).

Smith, F. and Jellema, A. (2016) *Introducing the Agriculture Open Data Package. BETA Version*. Global Open Data for Agriculture and Nutrition (GODAN), Wallingford, UK.

Smith, F., Fawcett, J. and Musker, R. (2017) Donor open data policy and practice: an analysis of five agriculture programmes [version 1; not peer reviewed]. *F1000Research* 2017 6, 1900 (document).

Sonka, S. (2014) Big data and the ag sector: More than lots of numbers. *International Food and Agribusiness Management Review* 17(1), 1–20.

Springer Nature (2016) *Over 600 Springer Nature journals commit to new data sharing policies*. Available at: https://group.springernature.com/gp/group/media/press-releases/archive-2016/over-600-springer-nature-journals-commit-to-new-data-sharing-policies/12000254 (accessed 8 March 2018).

Sundmaeker, H., Verdouw, C., Wolfert, S. and Freire, L.P. (2016) Internet of Food and Farm 2020. In: *Digitising the Industry – Internet of Things Connecting Physical, Digital and Virtual Worlds*. River Publishers, Gistrup/Delft, pp. 129–151.

USAID (2013) *Crowdsourcing Applications for Sustainable Development in Africa*. United States Agency for International Development (USAID), Washington, DC. Available at: http://pdf.usaid.gov/pdf_docs/PA00J7P7.pdf (accessed 8 March 2018).

Van Emon, J.M. (2016) The Omics Revolution in Agricultural Research. *Journal of Agricultural and Food Chemistry* 64(1), 36–44.

Walker, D., Kurth, T., Van Wyck, J. and Tilney, M. (2016) *Lessons from the Frontlines of the Agtech Revolution.* October 25. Boston Consulting Group (BCG), Boston, Massachusetts. Available at: https://www.bcg.com/publications/2016/process-industries-building-materials-strategy-lessons-frontlines-agtech-revolution.aspx (accessed 7 March 2018).

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M. *et al*. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, article 160018. Available at: https://www.nature.com/articles/sdata201618 (accessed 28 September 2018).

Wolfert, S., Ge, L., Verdouw, C. and Bogaardt, M.J. (2017) Big data in smart farming – a reviews. *Agricultural Systems* 153, 69–80.

Woodard, J., Andriessen, M., Cohen, C., Cox, C., Fritz, S. *et al*. (2017) Using ICT for Remote Sensing, Crowdsourcing, and Big Data to Unlock the Potential of Agricultural Data. Module 15. In: *ICT in Agriculture (Updated Edition): Connecting Smallholders to Knowledge, Networks, and Institutions.* World Bank, Washington, DC.