Congressional Research Service
Informing the legislative debate since 1914

Big Data in U.S. Agriculture

Megan Stubbs
Specialist in Agricultural Conservation and Natural Resources Policy

January 6, 2016

Congressional Research Service
7-5700
www.crs.gov
R44331

Summary

Recent media and industry reports have employed the term *big data* as a key to the future of increased food production and sustainable agriculture. A recent hearing on the private elements of big data in agriculture suggests that Congress too is interested in potential opportunities and challenges big data may hold. While there appears to be great interest, the subject of big data is complex and often misunderstood, especially within the context of agriculture.

There is no commonly accepted definition of the term *big data*. It is often used to describe a modern trend in which the combination of technology and advanced analytics creates a new way of processing information that is more useful and timely. In other words, big data is just as much about new methods for processing data as about the data themselves. It is dynamic, and when analyzed can provide a useful tool in a decisionmaking process. Most see big data in agriculture at the end use point, where farmers use precision tools to potentially create positive results like increased yields, reduced inputs, or greater sustainability. While this is certainly the more intriguing part of the discussion, it is but one aspect and does not necessarily represent a complete picture.

Both private and public big data play a key role in the use of technology and analytics that drive a producer's evidence-based decisions. Public-level big data represent records collected, maintained, and analyzed through publicly funded sources, specifically by federal agencies (e.g., farm program participant records and weather data). Private big data represent records generated at the production level and originate with the farmer or rancher (e.g., yield, soil analysis, irrigation levels, livestock movement, and grazing rates). While discussed separately in this report, public and private big data are typically combined to create a more complete picture of an agricultural operation and therefore better decisionmaking tools.

Big data may significantly affect many aspects of the agricultural industry, although the full extent and nature of its eventual impacts remain uncertain. Many observers predict that the growth of big data will bring positive benefits through enhanced production, resource efficiency, and improved adaptation to climate change. While lauded for its potentially revolutionary applications, big data is not without issues. From a policy perspective, issues related to big data involve nearly every stage of its existence, including its collection (how it is captured), management (how it is stored and managed), and use (how it is analyzed and used). It is still unclear how big data will progress within agriculture due to technical and policy challenges, such as privacy and security, for producers and policymakers. As Congress follows the issue a number of questions may arise, including a principal one—what is the federal role?

Contents

Contacts

The term *big data* continues to be a buzzword in many industries. While frequently discussed, no commonly accepted definition of the term *big data* exists.[1] In many cases, the terminology used to discuss the topic is not always consistent and can vary by industry and user. Based on the broad scope of the subject, this report will rely on a broad and general definition of big data compiled from resources within the agricultural community. The analysis also uses a terminology based on some of the more commonly used terms by government, research, and industry within the context of agriculture.[2]

While technology is a key and underlying component of big data use in the industry, its ongoing growth and evolution make it difficult to discuss in great detail. For the purposes of this report, technology—including hardware, software, and telemetry—and analytics are discussed in broad terms, using examples where relevant. These should not be considered exhaustive and do not imply endorsement of a named product or company. This report is also limited to the agricultural industry. Within the industry, farming and ranching operations use big data in different ways.[3] Where possible both are discussed; however, an increasing focus is placed on the use of big data in the production of crops. This imbalance is not intentional, but rather reflects the availability of resources for this discussion.

For the purposes of this report, the term *big data* will be discussed in two contexts: public and private.

- **Public-level big data** represent records that are collected, maintained, and analyzed through publicly funded sources, specifically by federal agencies (e.g., farm program participant records, Soil Survey, and weather data).
- **Private big data** represent records generated at the production level and originate with the farmer or rancher (e.g., yield, soil analysis, irrigation levels, livestock movement, and grazing rates).[4]

Both private and public big data play a key role in the use of technology and analytics that drive a producer's evidence-based decisions. Both also present challenges, such as privacy and security, for producers and policymakers.

Definition

As previously stated, a commonly accepted definition of the term *big data* does not exist.[5] At first glance, it appears that the term is used to describe a large collection of records. This categorization, however, is generally considered to be an understatement. For something to fall

[1] Lisa Arthur, "What Is Big Data?," *Forbes*, August 15, 2013, http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/.

[2] The terminology used in this report comes from several sources. Some are the author's definitions and terms, distilled from reading, listening, and researching previous articles, debates, and forums on the subject within the context of agriculture.

[3] Farmers and ranchers are collectively referred to as producers throughout this report.

[4] Other private big data sources exist that could have an impact or a usefulness in production agriculture (e.g., consumer purchasing information, origin tracking, and input markets). However, due to time constraints, they are not discussed in detail in this report. The focus is on production agriculture rather than other points in the agriculture value chain, including private big data created by agribusinesses.

[5] Kenneth Cukier and Viktor Mayer-Schoenberger, "The Rise of Big Data: How It's Changing the Way We Think About the World," *Foreign Affairs*, vol. 92, no. 1 (January/February 2013), p. 28.

into the category of big data it need not be *big*.[6] Rather, the term *big data* is often used to describe a modern trend in which the combination of technology and advanced analytics creates a new way of processing information that is more useful and timely. In other words, big data is just as much about new methods for processing data as about the data themselves.[7] Big data is viewed as dynamic and when analyzed can provide a useful tool in a decisionmaking process.

In the context of agricultural production, big data generally refers to the use of technology and advanced analytics for processing data in a useful and timely way. Big data may significantly affect many aspects of the agricultural industry, although the full extent and nature of its eventual impacts remain uncertain. Many observers predict that the growth of big data will bring positive benefits through enhanced production, resource efficiency, and improved adaptation to climate change.[8] While lauded for its potentially revolutionary applications, big data is not without issues. It is still unclear how big data will progress within agriculture due to challenges associated with both technical and policy issues.

Background

The use of technology in agriculture has continued to grow since the early part of the 20th century, when the industry shifted from the horse-drawn plow to mechanized tractors. The advent of plant genetics, chemical inputs, and, more recently, guidance systems has transformed the industry into one that is increasingly technology-intense and data-rich. The ability to generate, capture, and store data in the agricultural industry has continued to grow with the use of mobile technology and data management software. Additionally, external data sets are now readily available to the industry, allowing for a more complete picture of the world in which production agriculture occurs.

The technological advances that make up the modern computing environment have contributed to debate about big data. While data collecting is not new, especially in the context of public data collection, only since the advent of more efficient, mobile technologies and the digitization of data have large records been able to be evaluated and analyzed in a timely and more useful way. One key hallmark of big data is that it requires the use of analytical tools to extract value from it.[9] Without analysis, large quantities of data can be expensive, time consuming, and distracting.

From a policy perspective, issues related to big data involve nearly every stage of its existence, including its collection (how it is captured), management (how it is stored and managed), and use (how it is analyzed and used). These three stages exist in both public and private data, and are discussed in greater detail in the private big data section below.

Both private and public big data play a key role in the use of technology and analytics that drive a producer's evidence-based decisions. While discussed separately in this report, they are typically combined to create a more complete picture of an operation and therefore better decisionmaking

[6] Gary King, "Preface: Big Data Is Not About the Data!," in *Computational Social Science: Discovery and Prediction*, ed. R. Michael Alvarez (Cambridge: Cambridge University Press, 2015), http://j.mp/1PP466V.

[7] Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values*, May 2014, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf.

[8] For example, see USDA 2015 Agricultural Outlook Forum - Smart Agriculture in the 21st Century, "Plenary Panel - A Discussion on Innovation, Biotechnology, and Big Data," Arlington, VA, February 19, 2015, http://www.usda.gov/oce/forum/2015_Speeches/PanelDiscussionTranscription.pdf.

[9] Gary King, "Preface: Big Data Is Not About the Data!," in *Computational Social Science: Discovery and Prediction*, ed. R. Michael Alvarez (Cambridge: Cambridge University Press, 2015), http://j.mp/1PP466V.

tools. For example, companies that offer private big data products (discussed in the "Data Uses" section below) will combine agronomic, environmental, and operational data from multiple sources (i.e., public and private) in order to better describe current conditions and predict future results.[10] This combination is one of the key reasons big data for agriculture is viewed as such a valuable tool.

Congressional Action

On October 22, 2015, the House Agriculture Committee conducted a hearing on private big data in the agriculture industry.[11] Panelists discussed benefits and concerns related to private big data. Data ownership and privacy were chief among concerns, but most panelists agreed that little to no government intervention was desired.

No bills have been introduced in the last two Congresses relating specifically to big data in agriculture. Several bills in the 114th Congress could address issues that are potentially relevant to big data applications in agriculture, such as information sharing in cybersecurity, privacy, and notification of data breaches.[12]

Public Big Data

Public agricultural data sets are traditionally created through the use of surveys, samples, and statistical analysis. Advances in technology and analytics through big data have expanded this traditional role but also highlight another, previously less-used source—administrative data. While some public data records created by traditional means (e.g., surveys) are statutorily required to meet the mission of an agency, administrative data records are generally byproducts of program administration.[13] Whether the information is voluntarily or mandatorily collected also varies.

Using big data to inform federal actions is of increasing interest to policymakers. Specifically, it is the use of administrative data that raises the possibility of analyzing existing data records in order to make more efficient and better-informed decisions about federal farm programs and activities. It could also provide additional insight into behavioral and societal aspects of U.S. agriculture that might currently be underexplored.

Key Players

The agricultural industry has a number of publicly generated data sources.[14] Primarily, these sources are located at the U.S. Department of Agriculture (USDA). Other federal agencies, such

[10] For example, see services offered by companies like Monsanto's Climate Corporation: https://www.climate.com/; AGCO: http://www.agcocorp.com/brands/fuse-technologies.html; John Deere: https://www.deere.com/en_US/products/equipment/ag_management_solutions/ag_management_solutions.page?; and Pioneer: https://www.pioneer.com/home/site/us/programs-services/encirca-services/.

[11] http://agriculture.house.gov/news/documentsingle.aspx?DocumentID=2988.

[12] For more information, see CRS Report R43831, *Cybersecurity Issues and Challenges: In Brief* and related reports.

[13] Some scholars limit big data to "non-survey" data, others view it as large data sets that require "on-site access," and some include administrative data sets in the definition. The inclusion or exclusion of one data set or another varies based on the researcher's definition of big data. Because this report uses a very broad definition of big data, all of these areas are included. For additional discussion, see Council of Professional Associations on Federal Statistics, *Big Data and Federal Statistics*, brief, March 1, 2013, http://www.copafs.org/UserFiles/file/QtrlySummary2013March01.pdf.

[14] Unlike other countries, the United States does not have only one statistical agency. Currently, 13 agencies have (continued...)

as the U.S. Bureau of Labor Statistics, the National Oceanic and Atmospheric Administration (NOAA), and the National Aeronautics and Space Administration (NASA), also produce data sets important to the industry (e.g., meteorological information and satellite imagery).

USDA is arguably the agricultural industry's largest collector, manager, and user of public big data. A number of agencies within USDA participate in one or more of these activities. Whether the agency is a collector, manager, or user of traditional data or administrative data can sometimes make identifying an agency's role in big data difficult. Depending on how big data is defined, most USDA agencies could be considered users, generators, or managers of big data in one way or another.

Traditional Data

Traditional data is identified as the collection, management, and use of data obtained and analyzed through traditional means, such as surveys and sample collection. Examples of agencies and activities that generate traditional data include

- **National Agricultural Statistics Service (NASS)**—collects, manages, and analyzes survey data through the Census of Agriculture;
- **Economic Research Service (ERS)**—collects, manages, and uses resource, production, and financial data through the Agricultural Resource Management (ARM) survey;
- **Agricultural Research Service (ARS)**—collects, manages, and uses scientific data related to agriculture through its mission of research and information access;
- **Natural Resources Conservation Service (NRCS)**—collects, manages, and uses soil, water, and geospatial data through the Soil Survey program;
- **Agricultural Marketing Service (AMS)—**collects, manages, and uses price and sales information through its market news programs; and
- **World Agricultural Outlook Board (WAOB)**—analyzes commodity and market data to develop the World Agricultural Supply and Demand Estimates (WASDE) report.

Administrative Data

Some of these and other agencies have the capability of generating administrative data, generally as a by-product of program administration. While these agencies may not typically be considered big data agencies—their data is not generally made public and their data is not always aggregated or analyzed the same way as other big data—they nonetheless have similar issues as those of other big data agencies (i.e., security, privacy, technology capacity, and funding). Examples of these agencies include

- **Risk Management Agency (RMA)**—collects, manages, and uses individual yield and loss information to administer the Federal Crop Insurance program;

(...continued)

statistics as their primary focus. The Office of Management and Budget (OMB) is identified as a coordinator and establishes standards through guidance and also by administering various statutes related to privacy and quality. See U.S. Government Accountability Office, *Federal Statistical System: Agencies Can Make Greater Use of Existing Data, But Continued Progress Is Needed on Access and Quality Issues*, GAO-12-54, February 2012, http://www.gao.gov/assets/590/588856.pdf.

- **Farm Service Agency (FSA)**—collects and manages individual producers' farm record data, federal payments, and loan information used in administering various farm programs; and
- **NRCS**—collects and manages conservation plans, geospatial data, and conservation program activities and payments.

Benefits

A number of benefits to the agricultural industry are generally associated with public big data, including but not limited to

- **Authority**—Much public big data is governed by various statutes and guidance documents that establish standards for quality.[15] The use of common standards is believed to make public data more statistically reliable than some private sources, and public big data is therefore typically viewed as a trusted, authoritative source.
- **Confidentiality**—Public big data is governed by statutes and guidance documents that establish requirements for privacy.[16] In many cases, agencies are able to anonymize data so as to protect individual producers' identities. While some privacy and confidentiality concerns may also present challenges for agencies, their big data sources are generally thought to be more transparent and regulated than private ones.
- **Equal Access**—Technology has not only impacted the ability to collect, store, and analyze data, but also makes it more transparent and publicly accessible. The use of the Internet and open data initiatives allows greater access to and use of public big data.
- **Long-Term Investment**—The collection and use of data, including within the field of agriculture, has a long history in the United States. Frequently, the greatest value in traditional public data sets is derived from their ability to provide a baseline or benchmark for the industry as well as trends over time. The use of administrative data might provide new benchmarks and trends for areas previously not measured.

Challenges

While many see public big data as a trusted and reliable source, it is not without its challenges, including but not limited to

- **Resources**—The collection, management, and analysis of data are complex, cumbersome, and frequently resource intensive. Reduced federal budgets and staffing levels, combined with a technology adoption lag, have impaired the ability of some federal agricultural agencies to collect, manage, or use big data to

---

[15] Examples include the Paperwork Reduction Act (44 U.S.C. §3501 et. seq.) and the Information Quality Act of 2000 (Section 515 of P.L. 106-554); see OMB's *Agency Information Quality Guidelines*, https://www.whitehouse.gov/omb/inforeg_agency_info_quality_links.

[16] Examples include the Privacy Act of 1974 (5 U.S.C. §552a) and the eGovernment Act of 2002 (44 U.S.C. §3601 et seq.; see also OMB, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*, M-03-22, Washington, DC, September 26, 2003, https://www.whitehouse.gov/omb/memoranda_m03-22).

its fullest potential. Supporters of public big data frequently request additional or sustained resource levels.

- **Incorporation**—Traditional public data collection is typically through the use of statistically valid surveys. Increasingly, the technology and analytics of big data have allowed surveys to expand and have also resulted in the creation of additional value in administrative data. This has called into question whether the traditional public data sources can or should incorporate new big data sources. Are traditional and administrative big data compatible? What could be lost or gained if the current use of proper sampling were to change, relative to participation-based data sets?
- **Availability**—While not considered 100% "open," most public data is available in large part because of technology (see "Equal Access" bullet above). These are typically limited to the more traditional big data sources; however, there is an increasing interest in making administrative big data open as well. In some cases, public data access is driven by statutory language prohibiting its release.[17]
- **Security**—A number of high-level data breaches have raised concerns about the government's ability to protect its own big data sources.[18] Under current law, all federal agencies have cybersecurity responsibilities relating to their own systems.[19] While no known breach has occurred at USDA in recent years, some are concerned about the department's ability to protect information from similar attacks.[20]

---

[17] For example, see section 1619 of the Food, Conservation, and Energy Act of 2008 (P.L. 110-246), which prohibits the release of site-specific information related to farm programs.

[18] Julie Hirschfield Davis, "Hacking of Government Computers Exposed 21.5 Million People," *New York Times*, July 9, 2015, http://www.nytimes.com/2015/07/10/us/office-of-personnel-management-hackers-got-data-of-millions.html?_r=0.

[19] For more information and background, see CRS Report R43831, *Cybersecurity Issues and Challenges: In Brief*.

[20] In 2007, USDA announced that it had inadvertently exposed online sensitive information, including names and social security numbers, in a publicly available database. In 2014, the Office of Inspector General's office issued a report citing that the department was not adequately overseeing the security and performance of its Universal Telecommunications Network (UTN), which serves as the data network backbone for its customers and agencies. See USDA Office of Inspector General, *Management and Security over USDA's Universal Telecommunications Network*, Audit Report 88501-0002-12, Washington, DC, July 24, 2014, http://www.usda.gov/oig/webdocs/88501-0002-12.pdf.

Defining the Federal Role

Without a consistent definition of big data it can be difficult to fully define the federal government's role.[21] This is further complicated by an abundance of challenges associated with the government's ability to keep pace with a subject that is moving so quickly. How much leadership should the federal government offer to the agricultural industry? When federal agencies use public big data to provide services, how much control might they exert over the uses? Separate but related are questions concerning the federal role in private big data. Generally speaking, USDA does not have authority to regulate private big data unless it is somehow a function of a USDA program.[22] This raises the question: Is there a role for USDA in the private big data debate? And if so, what would that role look like? Also, how and should the federal government assure access to public goods provided through big data without impinging on competition and private-sector innovation?[23] Is there a role in eliminating or creating regulatory barriers to growth?

Private Big Data

Similar to public big data, private big data refers to the combination of technology and analytics used to process data. The key differences are who generates the underlying data, where that data is generated, and the purposes for which it is generated. In this report, private big data is limited to private data sets generated on the farm or ranch, by the producer, for enhancing the operation.[24] There are a number of key players in the agricultural industry that make private big data possible through the use of technology (e.g., software and hardware) and advanced analytics (e.g., descriptive, predictive, and prescriptive).[25]

Of the different stages of private big data—collection, management, and use—most people are familiar with the end result, or use of, the big data itself. This is frequently because it is the easiest to understand. It is where the producer most interacts with data—after it has been collected, analyzed, and turned into a usable form. For example, big data is used to create prescriptive plans that include recommendations on seed and fertilizer application rates, soil analysis, and localized weather reports. Other stages in the big data cycle, such as the collection and management of data, can be more complex, and frequently hold the biggest challenges for the agricultural industry.

---

[21] For a similar discussion, see CRS Insight IN10345, *Policy Implications of the Internet of Things*.

[22] While not the subject of this report, there are other federal agencies playing a role in aspects of big data through regulations, standards, and guidance. For example, see Federal Trade Commission, *Internet of Things*, staff report, January 2015, https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf or National Science Foundation (NSF), *Critical Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)*, program solicitation, NSF 16-512, February 9, 2016, http://www.nsf.gov/pubs/2016/nsf16512/nsf16512.htm.

[23] Adam D. Thierer, "The Internet of Things and Wearable Technology: Addressing Privacy and Security Concerns Without Derailing Innovation," *Richmond Journal of Law and Technology*, vol. 21, no. 2 (June 2015), http://jolt.richmond.edu/v21i2/article6.pdf.

[24] Some would argue that "no farm-level dataset is sufficient to make use of 'big data' without aggregating with other farms." For the purposes of this report individual farm-level data is included in the discussion because it is frequently discussed by the agricultural community at large. See Brian E. Whitacre, Tyler B. Mark, and Terry W. Griffin, "How Connected Are Our Farms?," *Choices*, vol. 29, no. 3 (3rd Quarter 2014), http://www.choicesmagazine.org/choices-magazine/submitted-articles/how-connected-are-our-farms.

[25] Tom Davenport, "5 Essential Principals for Understanding Analytics," *Harvard Business Review* (October 21, 2015), https://hbr.org/2015/10/5-essential-principles-for-understanding-analytics#.

Key Players

The number and type of players involved with private agricultural big data are constantly evolving. For that reason, this report arranges the discussion by examples of players within each stage of the big data process—collection, management, and use. These stages do not operate independently of one another and are, in practice, fluid in nature. In some cases the same player is active in more than one stage, or in all stages, of the process. This section is not intended to serve as an exhaustive list, but rather a starting point to discuss who and what is involved in each stage of the big data process.

Data Collection

The actual collectors of private big data are, in most cases, producers, who collect big data as part of their normal day-to-day activities. The collection stage, however, refers less to who collects the data than to how it is collected. Frequently, data collection involves physical technology, such as sensors, imagery, drones, radar, and other technologies all working together to provide detailed information about soil content, weeds and pests, sunlight and shade, nutrient deficiencies, moisture, and other factors. Physical technology forms only part of the data collection process, though. The other part is the network through which the technology communicates, typically the Internet. This is generally referred to as the "Internet of Things" (IoT) —networks of objects that communicate with other objects and with computers through the Internet.[26] Both have led to questions related to ownership, privacy, and security, among others.

Data collection is an ever-expanding area of big data and includes a number of key players, including but not limited to

- **Equipment Manufacturers**—In many cases the manufacturers of traditional farm equipment (e.g., tractors, combines, and implements) are well positioned to expand into data collection technologies since, in many cases, the technology is an extension of the equipment already in use.
- **Chemical Companies and Applicators**—The application and use of nutrients and pesticides are increasingly complex. As producers look for ways to reduce cost and become more efficient, the application of additives is frequently an area for improvement. Chemical companies are playing an increasing role in the research and development of data collection tools and methods to improve application use.
- **Multi-Use Technologies**—The agricultural industry increasingly is finding value in technologies used by other industries. For example, some farmers in the dairy industry are exploring the use of radio frequency identification (RFID), which is more commonly used in the shipping and transportation industries, to track movement, production, feed, and disease outbreaks in herds.[27] Frequently these companies have an interest in multiple industries, one of which could be agriculture.[28]

[26] For additional information, see CRS Report R44227, *The Internet of Things: Frequently Asked Questions*.

[27] S. F. Wamba and Alson Wicks, "RFID Deployment and Use in the Dairy Industry: Current Issues and Future Research Directions," conference paper published at the IEEE International Symposium on Technology and Society, Piscataway, NJ, June 7, 2010, pp. 172-179.

[28] For example, geospatial data is important in the context of agricultural big data. Companies such as Trimble, a geospatial positioning and navigation company, service a range of industries, including agriculture. See (continued...)

Data Management

Private big data management generally covers the organization, administration, and governance of characteristically large volumes of data. The goal of data management is to ensure a high level of data quality and accessibility for the end user, and ultimately the next stage of data use through big data analytics (discussed in the next section). While most see storage as the key function of data management, it often also includes processing and security measures.[29]

Key players, including private data managers and data banks, continue to expand. Others, such as data cooperatives, are entering the mix as producer-owned solutions for small to mid-sized operations that otherwise might face steep market access points individually. In some cases these players collect and organize the data for a fee, while in others they capture the value of the data by acting as a broker to trade or sell the data. Examples of the players include

- **Producers**—In some cases the function of data storage and management is conducted at the farm-level.[30] Concerns about security, and the sensitive nature of the data itself, have resulted in producers opting to store their data locally rather than through a third party or in a cloud computing environment.[31]
- **Data Collectors**—In some cases the same companies that offer data collection services also offer data management services. They are generally affiliated with other agricultural products (e.g., equipment, seed, or chemicals).
- **Independent Agricultural Data Banks**—These are private companies that offer to store, organize, and, in some cases, analyze data for a fee. They are generally independent of a company that provides other goods or services to the agricultural industry. Additional services may also be included, such as weather information, location services, and benchmarking.[32]
- **Data Cooperatives**—These are producer-owned information cooperatives (co-ops) that store, aggregate, and exchange data for their members. Similar to commodity co-ops, data co-ops pool members' data to create economies of scale and generate additional value and negotiating position. Data is then anonymized before being sold to interested parties. In return, members receive a portion of the proceeds from the sale as well as other infrastructure benefits provided by the co-op (e.g., storage and management).[33]

(...continued)

http://www.trimble.com/agriculture/index.aspx.

[29] Ash Ashutosh, "Best Practices for Managing Big Data," *Forbes*, July 5, 2012, http://www.forbes.com/sites/ciocentral/2012/07/05/best-practices-for-managing-big-data/.

[30] Brian E. Whitacre, Tyler B. Mark, and Terry W. Griffin, "How Connected Are Our Farms?," *Choices*, vol. 29, no. 3 (3rd Quarter 2014), http://www.choicesmagazine.org/choices-magazine/submitted-articles/how-connected-are-our-farms.

[31] Cloud computing refers a network of remote servers hosted on the Internet that stores, manages, and processes data. For more information, see CRS Report R42887, *Overview and Issues for Implementation of the Federal Cloud Computing Initiative: Implications for Federal Information Technology Reform Management*.

[32] For example, see the Farmers Business Network, https://www.farmersbusinessnetwork.com/.

[33] For example, see the Grower Information Services Cooperative, https://www.gisc.coop/.

Data Uses

The final stage of the big data process is the use of the data itself. This stage is where the actual tools are created and the value of private big data occurs for producers. A combination of big data sources is generally analyzed and packaged into an easily understandable and useful product.[34] These can cover the spectrum of qualitative analytical products: descriptive products (e.g., those that provide a better or more advanced way of looking at an operation); prescriptive products (e.g., those that provide timely recommendations for operation improvement based on real-time and historical data); and predictive products (e.g., those that use current and historical data sets to forecast future events and returns). Depending on the size and complexity, this stage may encompass all three types of analytical products.

The primary players and users of private big data products are the farmers and ranchers themselves. Other interested parties, however, recognize the value of big data and the ways it can be used beyond improving an individual operation (e.g., retail, marketing, or environmental improvement). Examples of key players in the use of private big data include

- **Farmers and Ranchers**—The end user of private big data is frequently the farmer or rancher from which the data originated. The value is derived from a big data product that offers improved production (e.g., lower costs, increased yields, or reduced inputs).
- **Retailers**—These are the creators of the big data products themselves. They derive value by analyzing the data, packaging it into a useable and timely product, and selling it to the producer. In some cases these are parties from other stages (i.e., data collectors or data managers) or retailers of other products (e.g., equipment and seed companies) or both. Increasingly, retailers themselves might find value in big data or the products they create separate from the value derived from the producer.[35] It is generally these groups that advocate for big data to get bigger in order to create a greater value for themselves and consumers.
- **Industry Groups**—The terms and uses of big data are often confusing and not well defined. Producer concerns related to privacy, security, and ownership abound, leaving many producers to look toward national commodity and agricultural industry groups for guidance regarding licensing language and data contracts. These groups are frequently active in establishing standards and issuing guidance for members on how to navigate big data.[36]
- **Environmental Interests**—Economics is only one benefit from the use of private agricultural big data. The power of technology and information can also result in a positive environmental effect through reduced inputs (e.g., fertilizer, pesticides, and water) and efficiencies (e.g., reduced air emissions through

[34] This is referred to by various names—precision agriculture, smart agriculture, satellite farming, etc.

[35] For example, seed companies could speed up research and establish a track record for new seed varieties through access to big data on production and yield. Or big data could allow whole farms to become research plots and study sites for agricultural researchers, instead of sections thereof as is currently the practice. Just as Amazon.com recommends items you might like based on what other customers who bought the same item also bought, farmers who bought a specific variety of soybean could use a similar "users who bought this also like" system to identify the best corn hybrid to purchase. Grant Gerlock, "Farmers Worry About Sharing Big Data," *Harvest Public Media*, February 16, 2014, http://harvestpublicmedia.org/article/farmers-worry-about-sharing-big-data.

[36] For example, see the American Farm Bureau Federation, *Privacy and Security Principles for Farm Data*, May 5, 2015, http://www.fb.org/tmp/uploads/PrivacyAndSecurityPrinciplesForFarmData.pdf.

reduced tillage overlap). This has caused various environmental interests to pay attention to agricultural big data and in some cases become active players.[37]

Benefits

As the use of private big data continues to expand, the number and scope of benefits continue to grow. Examples of observed benefits include

- **Production Benefits**—The number of benefits created by private big data—from increases in yields and greater efficiencies to reduced costs, inputs, and farm risk—continues to grow. Most of these benefits are directly received by the producers themselves through increased profits; however, spillover effects from increased farm revenue and environmental sustainability are also possible.
- **Environmental Benefits**—Similar to production benefits, the environmental benefits of private big data continue to develop. On-farm environmental benefits, including improved soil quality and water availability, can further benefit production as well as the environment. The reduction of inputs, including pesticides, fertilizer, water, and energy, often results in off-farm benefits through improved water quality and biodiversity.[38]
- **New and Expanded Business Opportunities**—The rise of big data and the advances in technology have created a number of new business opportunities around its use. While much of the attention has been directed at the creation of these large, prescriptive, analytic tools, other businesses have developed around or been modified by one or more aspects of the farming operation. For example, in the custom applicator industry, the use of big data allows the third party (i.e., custom applicator) to invest in the technology in order to provide a more precise application of various inputs as a service. For instance, instead of producers investing in the equipment required for certain big data collection, they may contract with another party who can use the same equipment and networks to determine and apply required amounts of inputs.
- **Real Time**—Advances in technology have led to the real-time collection and processing of data and congruently to the ability to receive real-time decisionmaking tools.[39] This can further expedite decisionmaking and possibly lead to increases in automation.

Challenges

Private big data is constantly and rapidly changing. The complex nature of big data and the pace with which it is moving have created much confusion in the industry. This has led to a number of challenges for production agriculture and raises questions about whether the industry can manage

[37] For example, the Environmental Defense Fund, among others, collaborated with United Suppliers to create Sustain—a fertilizer efficiency program: https://www.edf.org/media/edf-launches-initiative-reduce-fertilizer-pollution-commodity-grain-crops.

[38] John Antle, Susan Capalbo, and Laurie Houston, "Using Big Data to Evaluate Agro-environmental Policies," *Choices*, vol. 30, no. 3 (3rd Quarter 2015).

[39] Katherine Noyes, "Cropping up on Every Farm: Big Data Technology," *Fortune*, May 30, 2015, http://fortune.com/2014/05/30/cropping-up-on-every-farm-big-data-technology/.

these challenges and whether there is a role for federal involvement. Examples of challenges include

- **Ownership**—This is a widely discussed concern within the agricultural community. Many believe that the owner of private big data is obvious; it is the producer from which it is collected. This is not always true, however, because of how the data moves through the different stages of collection, management, and use. For example, when data is analyzed and provided to the producer during the use stage as a decision tool for their operation, the producer would seemingly be the clear owner because they derive the most value from the data's use. It is not that simple, however, because that same data might also have value to a third party who is able to aggregate it and analyze it for a different purpose (e.g., a company might be interested in seed, yield, and input rates for determining future pricing of their product). Therefore the answer to the question of ownership generally lies in who owns and controls the value of the data. As previously discussed, raw, large data sets frequently hold little value to the end consumer. It is the combination of this data with that of quantitative analytics that creates the value. Other questions also include the following: who owns the secondary and tertiary uses of the data; can this ownership be limited or expanded, and in what way; who is the owner if the data is collected under a separate contract (e.g., custom harvesting or custom applicator); and are the ownership rights of the landowner different from the producer when they are not the same party; what are the options, if any, to not having or limiting data collection?
- **Privacy**—Agricultural production presents unique challenges related to localized competition for resources (e.g., access to land and water). In many cases neighbors are competing against one another for access to such resources.[40] The concern of privacy is one that most producers acknowledge in the context of big data. Information related to yields and performance can hold incredible value.[41] The independent nature of farming tends to drive privacy concerns. While most concerns are in respect to market competition, there are an increasing number of privacy concerns related to regulatory and non-agricultural interest groups coming into possession of a producer's data.[42]
- **Security**—Many producers are concerned about the security implications of big data getting bigger and requiring the use of more advanced networks. As previously discussed, concerns about security and privacy have resulted in producers opting to store their data locally rather than through a third party or in a cloud computing environment. In some ways, however, this reduces the value of the data because in most cases, the bigger the data set, the greater the additional value created.[43] With recent security breaches in both the public and private sectors it may be unreasonable to believe that all data is fully secure.

---

[40] Quentin Hardy, "Working the Land and the Data," *New York Times*, November 30, 2014, http://www.nytimes.com/2014/12/01/business/working-the-land-and-the-data.html?_r=1.

[41] Dan Bobkoff, "Seed by Seed, Acre by Acre, Big Data Is Taking over the Farm," *Business Insider*, September 15, 2015, http://www.businessinsider.com/big-data-and-farming-2015-8.

[42] For example, an American Farm Bureau Federation data privacy survey of its members found that over 77% of respondents were concerned that the farm data could get in the hands of an entity and be used for regulatory purposes. http://www.fb.org/tmp/uploads/AFBF_Final_Big_Data_Survey_Highlights_9-8-2014.pdf.

[43] Many researchers, scientists, and consultants advocate for combining all production agricultural data in order to draw (continued...)

- **Market**—Unbalanced access to information can frequently distort the marketplace. Some speculate that the use of yield or sales data could create a market advantage for some input companies (e.g., seed and fertilizer companies). Others are concerned that big data will increase the competitiveness of rental agreements,[44] while others see the potential for big data to affect commodity markets.[45]
- **Expense**—Similar to the discussion under public big data, the collection, management, and analysis of data are not free. The use of big data products can come at a cost, albeit one which larger operations are able to afford with the anticipation of higher returns.[46] Smaller operators, however, may face challenges related to economies of scale given the high price of equipment and services related to big data.[47] Arguments have also been made that while some private big data products offer greater returns with greater investment, some smaller companies present small operators with options to capture the benefits of big data through more affordable technologies and services.[48]
- **Infrastructure**—Recent surveys and studies indicate that, in general, rural areas tend to lag behind urban and suburban areas in broadband deployment. In the context of big data, the ability to send and receive data as well as the speed with which this occurs is central to its use and adoption.[49] The federal government has played a role in addressing rural broadband infrastructure through financial assistance programs and spectrum policy for wireless connectivity.[50]

(...continued)

new and important conclusions, seeing this as the pinnacle of big data. John Antle, Susan Capalbo, and Laurie Houston, "Using Big Data to Evaluate Agro-environmental Policies," *Choices*, vol. 30, no. 3 (3rd Quarter 2015) and Grant Gerlock, "Farmers Worry About Sharing Big Data," *Harvest Public Media*, February 16, 2014, http://harvestpublicmedia.org/article/farmers-worry-about-sharing-big-data.

[44] For example, if a neighboring farmer acquired production and yield information for a specific plot of land, it might spur unwanted competition to rent the land, creating higher rental rates.

[45] For example, some are concerned that commodity traders could use the data to make bets on futures contracts. If this lowers futures-contract prices early in the growing season, it might squeeze the profits that farmers otherwise could lock in for their crops by selling futures. Jacob Bunge, "Big Data Comes to the Farm, Sowing Mistrust," *Wall Street Journal*, February 25, 2014, http://www.wsj.com/articles/SB10001424052702304450904579369283869192124.

[46] For example, private big data products are reported to improve the average corn harvest to more than 200 bushels an acre from the current 160 bushels. This gain could generate an extra $182 an acre in revenue, based on 2014 prices, potentially a large amount of revenue if multiplied by thousands of acres on a large operation. Source: Rich Cooper, *The Seeds of Innovation – Big Data Reshaping U.S. Agriculture*, U.S. Chamber of Commerce Foundation, March 13, 2014, http://www.uschamberfoundation.org/blog/post/seeds-innovation-big-data-reshaping-us-agriculture/34140.

[47] Michael Hickins, "Op-Ed: For Small Farmers, Big Data Adds Modern Problems to Ancient Ones," *Wall Street Journal*, February 25, 2014, http://www.luxresearchinc.com/sites/default/files/WSJ_2-25-14.pdf.

[48] For example, Farmlogs utilizes smartphone-type software to schedule farm tasks and analyze yields. See https://farmlogs.com/; source: Quentin Hardy, "A Low-Cost Alternative to Pricy Big Data on the Farm," *New York Times*, December 1, 2014, http://bits.blogs.nytimes.com/2014/12/01/a-low-cost-alternative-to-pricy-big-data-on-the-farm/?_r=0.

[49] Brian E. Whitacre, Tyler B. Mark, and Terry W. Griffin, "How Connected Are Our Farms?," *Choices*, vol. 29, no. 3 (3rd Quarter 2014), http://www.choicesmagazine.org/choices-magazine/submitted-articles/how-connected-are-our-farms.

[50] For additional discussion, see CRS Report RL30719, *Broadband Internet Access and the Digital Divide: Federal Assistance Programs*, and CRS Report RL33816, *Broadband Loan and Grant Programs in the USDA's Rural Utilities Service*.

- **Technology System Failure**—Users of technology know that it can be wonderful until it stops working. Basic system failures and limitations will continue to be a challenge for big data. In some cases the collection of the wrong data or inaccurate data can lead to poor decisions, rather than improved ones.
- **Technology Adoption**—The adoption of technology itself could prove a hurdle for the agricultural industry. While some producers are open to the adoption of technology on their operations, the sector's aging producer population may be slow to adopt it.

Concluding Remarks

Both public and private big data can exist independently of one another; consequently, they are discussed separately in this report. This was done by design to illustrate an increasingly complicated topic. It does not, however, represent how observers think about big data in the context of agriculture. Most see big data in agriculture at the end use point, where farmers use precision tools to potentially create positive results like increased yields, reduced inputs, or greater sustainability. While this is certainly the more intriguing part of the discussion, it is but one aspect and does not necessarily represent a complete picture.

Big data is a complicated topic, not only from a technological and analytical standpoint, but also from a legal, ethical, and regulatory standpoint. The number of key players continues to grow, as does the list of benefits and challenges. As Congress follows the issue a number of questions may arise, including a principal one—what is the federal role?

Author Contact Information

Megan Stubbs
Specialist in Agricultural Conservation and Natural Resources Policy
mstubbs@crs.loc.gov, 7-8707