Agricultural Data Innovation:
Implications for the USDA

Douglas Lawrence and William A. Salas

April 2019
About the Authors

**Douglas Lawrence**, a North Dakota native, worked for the Natural Resources Conservation Service (NRCS) of the U.S. Department of Agriculture (USDA) for 37 years in various capacities. He retired as the NRCS Deputy Chief for Soil Survey and Resource Assessment. In this position, he was responsible for overseeing the world’s leading soil survey program as well as two key natural resource assessment activities: the National Resources Inventory and the Conservation Effects Assessment Project. His responsibilities also included directing the NRCS’s international conservation activities. In 2006, Dr. Lawrence was co-recipient of a USDA Honor Award for his work on energy conservation; in 2011, he was co-recipient of another USDA Honor Award related to reorganizing the soil survey program; and in 2008 he received the NRCS Chief’s Circle of Excellence award for his work on the 2008 Farm Bill. Dr. Lawrence holds a PhD in economics from the University of New Hampshire; a master’s degree in agricultural and resource economics from the University of Maryland; and a bachelor’s degree in economics from the University of Wisconsin.

**William Salas** is President and Chief Scientist of Applied GeoSolutions (AGS). Prior to forming AGS in 2000, Dr. Salas worked for five years at the Jet Propulsion Laboratory of the National Aeronautics and Space Administration and for 10 years at the Complex Systems Research Center at the University of New Hampshire (UNH). Dr. Salas received his BS in mathematics and physics at the University of Vermont and his MS and PhD in natural resources at UNH. Dr. Salas’s expertise and interests include the use of remote sensing tools to model land use and land cover change, soil biogeochemical modeling, and geographic information system tools for environmental applications. Dr. Salas is an active member of many panels, including the Scientific Advisory Panel for the ALOS Kyoto and Carbon Initiative under the auspices of the Japanese space agency NASDA. He has also served on the Global Observation of Forest Cover’s High Resolution Design Team and coauthored the final design document on the role of high-resolution optical and radar systems for operational monitoring of global forests.
AGree provides a forum for discussion about timely issues affecting agricultural producers across the country. We support innovative agricultural policies and programs that reduce risk, increase producer profitability, and improve environmental outcomes on U.S. working lands. Our diverse group of producers, former USDA leaders, researchers, academics, and representatives from environmental and conservation NGOs sees value in harnessing the power of agriculture data to inform producer decision-making while protecting data privacy.

Thanks to new technological advancements, agriculture data collection practices are growing rapidly in the private and public sector. These innovations have the potential to vastly improve the way producers make land management decisions, by providing them with data-driven insights about which practices reduce risk and improve yield. Private companies have perfected this model and are creating new markets for “smart agriculture.” There are opportunities for the U.S. Department of Agriculture (USDA) to improve its data management processes to better meet the needs of U.S. farmers and ranchers.

This paper was written by Douglas Lawrence (former NRCS Deputy Chief of Soil Survey and Resource Assessment) and William Salas (President of Applied GeoSolutions). It explores a range of agriculture data collection trends and their implications for USDA and the private sector. Importantly, the paper proposes recommendations for data collection and analysis methods that can be used by USDA to generate data that is more timely, efficient, accessible, and robust.

While the concepts discussed in this paper will enrich AGree and others’ discussions, they do not represent official AGree positions.

Deborah Atwood
Executive Director, AGree
## Contents

**Introduction** ...................................................... 1

**USDA Data Collection.** ........................................... 1

  - Scope of USDA Data Collection ........................................... 1
  - Strengths of USDA Data .................................................... 2
  - Current Data Collection Methods ............................................ 2
  - 2018 Farm Bill ............................................................. 4

**Private-Sector Investment in Agriculture Data Collection** .................. 4

**Issues** ............................................................ 4

  - Privacy ................................................................... 4
  - Declining Response Rates to USDA Surveys ............................ 5
  - Cost of Adoption ........................................................... 6
  - Access .................................................................... 6
  - Security .................................................................. 7
  - Private Data Competition .................................................... 7

**Advances in Agricultural Conservation Data Collection** ...................... 7

  - Ag Tech Investments in Data Management and Information Systems ........ 7
  - Ag Tech Investments in Disruptive Data Management Paradigms ................. 8
  - Centralized and Integrated Federal Data Collection .............................. 8
  - Public–Private Data Collection: Remote Sensing Partnerships .................. 8

**Recommendations** ................................................ 10

  1. Ensure Data Privacy ...................................................... 10
  2. Identify USDA Data and Analytical Needs ............................... 11
  3. Develop Data Partnerships with the Private Sector ...................... 11
  4. Address Declining Survey Response Rates ................................... 11
  5. Improve Interagency Coordination of USDA Data Collection .............. 11
  6. Create a Data Warehouse .................................................. 12
  7. Fund Innovative Data Collection, Analysis, and Access Initiatives ........ 12

**Reference List** .................................................... 12

**Endnotes** ........................................................ 13
Introduction

The collection of farm and ranch production data—especially in the private sector—is growing rapidly, driven largely by technological advancements. In 2016, the global market for “smart agriculture” technologies and services was estimated to be $5.1 billion. Another estimate projects that the global “big data” market in agriculture will grow by 20.1 percent per year from 2018 to 2022. These rapid advances in data collection and powerful analytical tools, including artificial intelligence and machine learning, are driving increased production efficiency, farm and ranch profitability, and reduced environmental impacts.

Innovations in agricultural data collection and analysis offer opportunities for the U.S. Department of Agriculture (USDA) to create more effective programs and reduce costs. A broad range of stakeholders could benefit from improvements in the USDA’s data collection, analysis, and warehousing methods. Such improvements could reduce the “survey response burden” felt by farmers and ranchers, for example, saving them valuable time and potentially also improving survey response rates. Improvements could also reduce the cost of USDA programs and increase their effectiveness, as well as help the people who need the information—farmers and ranchers—improve productivity, profitability, and environmental outcomes.

Farm production surveys were introduced in 1863 to help ensure that farmers and ranchers had free access to market-related information, in order to reduce market information asymmetries that allowed mills, granaries, slaughterhouses, and other food processing businesses to price agricultural commodities at below free-market values. The dramatic expansion of private agricultural data collection leads to new information asymmetries and potentially threatens existing USDA agricultural data collection and use. Improving the timeliness of USDA agricultural production reports will help producers and others keep up with production information generated by the private sector for sale to large commodity buyers and speculators.

This paper explores a range of agricultural data collection trends and their implications for the USDA and the private sector. The paper describes:

- USDA data collection, analysis, and warehousing related to agricultural production and conservation;
- Private-sector investment in agricultural conservation data collection, including satellite-based remote sensing; and
- Advances in agricultural conservation data and farm management information systems.

The paper concludes with recommendations for data collection and analysis methods that have the potential to be used by the USDA to generate data and information that are more timely, efficient, accessible, and robust.

USDA Data Collection

Scope of USDA Data Collection

The USDA has been the agriculture sector’s largest collector, manager, and user of agricultural data since the 1860s. USDA data collection can effectively be divided into two broad categories: traditional data and administrative data. Traditional data refers to the collection, analysis, and warehousing methods. Such improvements could reduce the “survey response burden” felt by farmers and ranchers, for example, saving them valuable time and potentially also improving survey response rates. Improvements could also reduce the cost of USDA programs and increase their effectiveness, as well as help the people who need the information—farmers and ranchers—improve productivity, profitability, and environmental outcomes.

The USDA has been the agriculture sector’s largest collector, manager, and user of agricultural data since the 1860s. USDA data collection can effectively be divided into two broad categories: traditional data and administrative data. Traditional data refers to the collection, analysis, and warehousing methods. Such improvements could reduce the “survey response burden” felt by farmers and ranchers, for example, saving them valuable time and potentially also improving survey response rates. Improvements could also reduce the cost of USDA programs and increase their effectiveness, as well as help the people who need the information—farmers and ranchers—improve productivity, profitability, and environmental outcomes.

The paper explores a range of agricultural data collection trends and their implications for the USDA and the private sector. The paper describes:

- USDA data collection, analysis, and warehousing related to agricultural production and conservation;
- Private-sector investment in agricultural conservation data collection, including satellite-based remote sensing; and
- Advances in agricultural conservation data and farm management information systems.

The paper concludes with recommendations for data collection and analysis methods that have the potential to be used by the USDA to generate data and information that are more timely, efficient, accessible, and robust.

USDA Data Collection

Scope of USDA Data Collection

The USDA has been the agriculture sector’s largest collector, manager, and user of agricultural data since the 1860s. USDA data collection can effectively be divided into two broad categories: traditional data and administrative data. Traditional data refers to the collection, analysis, and warehousing methods. Such improvements could reduce the “survey response burden” felt by farmers and ranchers, for example, saving them valuable time and potentially also improving survey response rates. Improvements could also reduce the cost of USDA programs and increase their effectiveness, as well as help the people who need the information—farmers and ranchers—improve productivity, profitability, and environmental outcomes.

The USDA has been the agriculture sector’s largest collector, manager, and user of agricultural data since the 1860s. USDA data collection can effectively be divided into two broad categories: traditional data and administrative data. Traditional data refers to the collection, analysis, and warehousing methods. Such improvements could reduce the “survey response burden” felt by farmers and ranchers, for example, saving them valuable time and potentially also improving survey response rates. Improvements could also reduce the cost of USDA programs and increase their effectiveness, as well as help the people who need the information—farmers and ranchers—improve productivity, profitability, and environmental outcomes.

USDA Data Collection

Scope of USDA Data Collection

The USDA has been the agriculture sector’s largest collector, manager, and user of agricultural data since the 1860s. USDA data collection can effectively be divided into two broad categories: traditional data and administrative data. Traditional data refers to the collection, analysis, and warehousing methods. Such improvements could reduce the “survey response burden” felt by farmers and ranchers, for example, saving them valuable time and potentially also improving survey response rates. Improvements could also reduce the cost of USDA programs and increase their effectiveness, as well as help the people who need the information—farmers and ranchers—improve productivity, profitability, and environmental outcomes.
• The ERS manages and analyzes resource, production, and financial data collected by the Agricultural Resource Management Survey, and it analyzes data from numerous other surveys.

• The ARS collects, manages, and uses scientific data related to agriculture through its mission of research and information access.

• The NRCS collects, manages, and uses soil, water, and geospatial data collected through the soil survey program, the National Resources Inventory (conducted every five years), the snow survey, and the Conservation Effects Assessment Project. USDA agencies also collect administrative data, generally as a byproduct of program administration. Although administrative data is not sampled like traditional data, it faces the same issues related to security and privacy. Examples of agencies that collect administrative or program data include the Risk Management Agency (RMA), the Farm Service Agency (FSA), and the NRCS.

• The RMA collects, manages, and uses individual yield and loss information to administer the Federal Crop Insurance program.

• The FSA collects and manages data related to farm production, the Conservation Reserve Program, field boundaries, federal commodity program payments, crop and livestock disaster payments, and loan information used in administering various farm programs.

• The NRCS collects and manages data associated with conservation planning, conservation program contracts, program payments, and conservation easement management.

**Strengths of USDA Data**

The USDA has a long track record of collecting and managing data in ways that help to ensure quality and confidentiality and allow for the analysis of long-term trends (Stubbs 2016).

**Quality:** Much of the data collected by the USDA is governed by statutes and guidance documents that establish standards for quality. 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:** USDA data are protected by statutes and guidance documents that establish requirements for privacy. In many cases, agencies are able to anonymize data to protect individual producers’ identities. While some privacy and confidentiality concerns may present challenges for agencies, their big data sources are generally thought to be more transparent and regulated than private sources of agricultural data.

**Long-Term Commitment:** The greatest value in traditional public data sets is derived from their ability to provide baselines, benchmarks, and time series data that can be used to evaluate agricultural and natural resources trends over time.

**Current Data Collection Methods**

The USDA currently uses the following data collection methods.

**Farmer Surveys**

• **Multimodal Data Collection:** Much of the data collected by NASS is gathered through surveys of producers. NASS employs several strategies to collect this data, including telephone surveys, mailed paper surveys, surveys filled out with the help of enumerators, and web-based surveys.

• **Survey and Administrative Data Integration:** Integrating survey data with program data can reduce the burden on farmers and potentially be more efficient. However, program implementation agencies often do not have the legislative authority to collect data beyond what they need to administer their programs. Agencies face pressure to streamline administrative data collection to reduce the burden on producers.
Soil Survey

NRCS soil survey data, including tables and geographic information system (GIS) data layers, can be viewed or downloaded at the USDA’s Geospatial Data Gateway available on the Soil Survey Geographic database. The soils information was gathered by site visits and soil samples that were analyzed in laboratories. The mapping is intended for natural resource planning and management by landowners, managers, townships, and counties, among others.

Remote Sensing

- **Aerial Imagery**: Aerial photography is an important data collection method used by the USDA. The FSA, through its Aerial Photography Field Office, maintains an enormous aerial film library that includes more than 10 million images dating back to 1955. Historic aerial images play a vital role in environmental assessments, change detection, and property boundary disputes. The FSA’s National Agriculture Imagery Program provides digital imagery to the USDA Service Center agencies that utilize GIS to administer federal farm and conservation programs.

- **Satellite Imagery**: Agricultural data collection through satellite remote sensing has been a mainstay of the USDA, the U.S. Geological Survey (USGS), and other federal agencies for well over a decade. The Multi-Resolution Land Characteristics consortium, a group of 10 federal agencies, has been creating the National Land Cover Database every five years since 2001. Separately, the USGS’s Land Cover Trends project includes land use and land cover data from 1972 through 2000. Both of these initiatives use multi-temporal Landsat satellite imagery to map agricultural land at field scales. At NASS, the Cropland Data Layer (CDL) Program utilizes remote sensing to provide in-season crop acreage estimates. While the CDL Program has been in existence since 1997, it was primarily a research effort until 2009, when it became an operational program.

- **LiDAR Imagery**: The NRCS partners with other federal agencies, states, and counties to acquire Light Detection and Ranging (LiDAR) data about surface elevation. National elevation data could be used by the NRCS to improve the planning and design of conservation systems for farmers and ranchers as well as to update soil maps to a level of accuracy compatible with precision agriculture.

USDA Agricultural Conservation Data Collection

Data regarding the status of conservation practices on agricultural land has been collected by the NRCS primarily in two ways: (1) Large-scale, statistically designed sampling projects, including the National Resources Inventory (NRI) and the Conservation Effects Assessment Project (CEAP) cropland surveys, and (2) less-structured assessments such as the crop residue management survey conducted by the Conservation Technology Information Center (CTIC).

- Data for the NRI survey are collected annually, and reports are published by the NRCS every five years. The NRI includes information on land use, soil erosion rates, and the adoption of conservation practices.

- The CEAP cropland survey regarding the adoption of conservation practices has been conducted twice (2003/4 and 2015/16) and consists of detailed interviews with farmers that cover three years of management information as well as material gathered from NRCS program records.

- From 1989 to 2004, the CTIC worked with the NRCS and local soil and water conservation districts to conduct a crop residue management survey of tillage practices across the country. A national residue management survey has not been conducted since 2004. However, statistics regarding cover crops at the local and regional scales have been generated since then via a cropland roadside transect survey (CTIC 2008).
Some states have been collecting state-level estimates of the adoption of conservation practices since the national survey efforts stopped in 2005. Indiana, for example, has been conducting transect surveys since the mid-1990s and now includes information on crop residue, tillage, and cover crop adoption. The CTIC, in collaboration with the USDA’s Sustainable Agriculture Research and Education program, has been conducting cover crop surveys from the 2012/13 cover crop season through the 2016/17 season. Recent advances in remote sensing are beginning to replace expensive, ground-based surveys regarding the adoption of conservation practices. That said, surveys will continue to be an important source of training data for machine learning algorithms that are used to analyze remotely sensed data.

2018 Farm Bill

On December 20, 2018, the Agriculture Improvement Act of 2018—also known as the 2018 Farm Bill—passed into law. Section 12618 of the Farm Bill directs the USDA to generate a report identifying available Departmental data on conservation practices and the effects of such practices on farm and ranch profitability, including effects relating to crop yields and soil health. The report will summarize the steps the U.S. Secretary of Agriculture will have to take to provide access to the data to university researchers, including technical, privacy, or administrative considerations; the safeguards linked to providing access to the data; appropriate procedures to maximize research benefits; and recommendations relating to federal authorizations needed to allow access to the data.

Private-Sector Investment in Agriculture Data Collection

According to AgFunder News, investment in agricultural technology reached $10.1 billion in 2017. The collection, analysis, and use of farm and ranch data by the private sector is enjoying extraordinary growth. In some cases, this growth is associated with the trend to manage production and conservation at the sub-field level using precision agriculture, which employs sophisticated data collection and complex analysis, integrated with innovative equipment, to radically improve farm management.

A good example of this is the new Land O’Lakes Truterra™ Insights Engine, a web-based “interactive on-farm digital platform that will help farmers advance their stewardship goals and return-on-investment in real time, acre-by-acre and help food companies measure sustainability progress.” Through this platform, field and sub-field data related to crop production, soils, conservation, and economics are entered and stored in WinField United’s Data Silo™, where it can be accessed and updated over time. The Truterra Insights Engine provides agricultural retailers and farmers with a wide array of information on conservation and profitability. Moreover, it allows the farmer to undertake “what-if” analyses, so they can calibrate the management of their crops and fields to optimize income while improving soil health and reducing externalities.

As the data collected and stored by the private sector grows, there may be opportunities for the USDA to collaborate on data collection for certain variables.

Issues

The most pressing issues relating to the USDA’s collection of data include privacy, declining response rates, cost of adoption, access, security, and competition with privately collected data.

Privacy

In general, farmers wish to control their data and have three main concerns about how the use of that data will impact the competitiveness of their businesses. First, a farmer’s production system is his or her intellectual property, and protecting that intellectual property is important to maintaining a competitive advantage. Second, farmers have concerns about whether their data will be used by service providers...
(e.g., ag retailers, seed companies, and agrochemical companies). Third, farmers are concerned that their data could be used to regulate their operations. The USDA has been sensitive to privacy concerns and has an excellent track record of protecting the privacy of farmers and ranchers’ data. For a detailed discussion of the data privacy issue, see the AGree paper entitled *USDA and Agricultural Data: Improving Productivity while Protecting Privacy* (Janzen and Ristino 2018).

In the private sector, the adoption of ag data technologies has been constrained by concerns about data privacy. Contracts and regulations regarding the use and ownership of data are essential. During the past few years, national agriculture groups have met with big data companies to establish data privacy principles that apply to agreements between farmers and technology providers. For example, groups such as the American Farm Bureau Federation and the National Corn Growers Association have worked with companies such as The Climate Corporation, John Deere, DuPont Pioneer, and others to develop principles for data ownership.

**Declining Response Rates to USDA Surveys**

Surveys of all types, public and private, are experiencing declining response rates, a challenge that is sometimes referred to as “survey fatigue.” A recent Congressional Research Service report (Schnepf 2017) indicated that NASS targets an 80 percent response rate in all of its surveys. That is an extraordinarily high response rate compared to customer surveys in the private sector, where response rates for email-delivered online surveys, for example, average about 14 percent. A given survey response rate will depend on a number of factors, including customer attitudes toward the entity conducting the survey, therefore response rates range widely across surveys.11

One study found that NASS response rates, after being in the 80 percent to 85 percent range in the early 1990s, have declined to around 60 percent (Johansson et al. 2017). The decline accelerated after 2010, raising concerns about a long-term trend toward lower response rates.

**Response Rates for NASS Acreage and Production Surveys Have Been Declining since the Early 1990s**

![Graph showing declining response rates from 1992 to 2016](Image)

NASS is actively addressing this issue, and NASS survey response rates are still generally much higher than the rates universities are able to achieve, perhaps in part because farmers and ranchers understand how their responses are used to benefit the agriculture sector. In addition, NASS uses multimodal survey approaches that include phone calls, mailings, and web-based data collection. NASS also uses a range of statistical techniques to help with lower response rates, such as reweighting survey data that has lower-than-desired response rates.

The issue of response rates can be more problematic at the county level, because of smaller sample sizes. This becomes particularly important in a program such as the Agriculture Loss Coverage-County (ARC-CO) commodity program, which is administered by the FSA and provides revenue loss coverage at the county level. Payments are issued when the actual county crop revenue of a covered commodity is less than the ARC-CO guarantee for that commodity. The FSA uses NASS county-level yields to compute the actual county crop revenue. NASS can only publish a county yield estimate if at least 30 producers return the yield survey data, or if at least three responses represent at least 25 percent of the total county acreage. When NASS county yield data are not available, the FSA uses the RMA crop insurance average yield for the county. However, RMA yield data are generally higher than NASS yield data, which means farmers in adjacent counties can potentially receive different ARC-CO payments. If neither NASS nor RMA yield data are available, FSA policy calls for the FSA state committee to determine county yields using “best available data.”

In 2010, NASS established a response rate review team to improve response rates, which resulted in the rates stabilizing, and, in some cases, improving. In addition to multimodal data collection, NASS fills data gaps by using previously reported data in addition to questionnaires, and the agency is working to develop techniques that use that previously reported data without introducing bias. Another approach to reducing survey fatigue is to use, where appropriate, administrative or program data from other agencies to supplement surveys. Program data, however, can create statistical issues because they are not drawn from a carefully designed sample and may not be representative of the underlying population. Here again, NASS is working on techniques to integrate program or administrative data in a way that is statistically defensible.

An alternative approach to dealing with survey fatigue is to allow producers to voluntarily opt-in to a selected survey. Because some producers see value in the surveys, they have an interest in making sure their data are included. However, this approach could have the unintended consequence of introducing bias in the sample.

Cost of Adoption

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

Access

Public access to data collected by the USDA is typically limited to the more traditional big-data sources, such as the Agricultural Census. Moreover, access to the raw data is limited to entities that have specific project-related approval from NASS. There is an increasing interest in making administrative big data open as well. In some cases, public data access is limited by statutory language prohibiting its release. The U.S. Office of Management and Budget recognizes that data is a valuable national resource and requires agencies to make data available, discoverable, and usable.
Security

A number of highly publicized data breaches have raised concerns about the government’s ability to protect its own big data sources. Under current law, all federal agencies have cybersecurity responsibilities relating to their own systems. While no known breach has occurred at the USDA in recent years, there are ongoing concerns about protecting USDA information.

Private Data Competition

When it comes to agricultural data, perhaps the prime area for the USDA and the private sector to compete is with regard to data associated with agricultural markets. Much of the data collected by NASS—including acres planted, estimated yields, livestock production, and a host of other surveys—is closely watched by commodity markets.  

The commodity production data that NASS publishes—along with the monthly World Agricultural Supply and Demand Estimates report prepared by the USDA’s World Agricultural Outlook Board—seeks to ensure that producers, businesses, investors, and others have equal access to U.S. and international crop supply, demand, and price projections for major U.S. program crops and livestock production activities.

At the same time, a large network of private-sector, fee-based agricultural market news and information services have developed since the early 1970s. These private information services have largely complemented the USDA’s commodity reporting. Most private-sector market information businesses use NASS and other USDA data to develop their own reports, which may include expert opinion from market analysts (Schnepf 2006).

However, companies such as Planet are offering commodity traders as well as farmers satellite-based data that provides near-real-time information on crop acreage, biomass production, and even crop health. In the near term, this is probably an area where the private sector can offer commodity traders a potential advantage over survey data collected by NASS.

Advances in Agricultural Conservation Data Collection

Both the private sector and the public sector are investing in new data management and information systems and creating new models for data analysis and systems. These developments are creating for new opportunities for public-private partnerships.

Ag Tech Investments in Data Management and Information Systems

Collecting data on fertilizer use, conservation practices (tillage and cover cropping), yield, and soil type is critical for enhancing knowledge of management impacts on soil health and how conservation practices can improve crop resilience. Once aggregated, these data can be used to develop strategies to minimize crop production risks relating to cropping systems, soils, and conservation practices. Private-sector investment in agricultural technologies has led to the development of a wide array of farm management information systems. These systems present an opportunity, given the collection of massive amounts of data that can be analyzed, to improve farmers’ bottom lines and benefit the environment. While farm management systems are valuable for on-farm use and decision making, there is little interoperability across platforms, which makes it difficult to aggregate the data to garner information at scale. However, farmers’ privacy concerns are beginning to be addressed and data are increasingly available in an aggregate form that benefits agriculture. Investments in the ag tech community are needed to address data connectivity across digital platforms, to scale the data from multiple platforms, and to develop cybersecurity systems to preserve confidential data. Efforts toward standardization and well-defined Application Programming Interfaces are facilitating the sharing and exchange of data across multiple platforms.
AgTech Investments in Disruptive Data Management Paradigms

Recent investments in new ag tech startups have the potential to disrupt the way data are handled and enable better identity preservation and aggregation of data to build the business for conservation agriculture. One of these new ventures is Farmers Business Network, which aggregates data from individual farmers and shares those data, anonymously, with other farmers in the network to let growers share insights.

A second example is Indigo Ag, which was formed in 2014 to provide microbially treated seeds to optimize crop yields and environmental performance. Indigo has expanded its business products to include farm management tools, agronomic services, and production contracts in which growers receive price premiums on crops grown with Indigo seeds. Indigo is raising additional capital to build out its marketplace to directly connect farmers with buyers. This direct connection may facilitate identity preservation related to farmer production practices, including the use of conservation practices, something that is currently not possible in the commodity supply chain.

Lastly, there is movement toward the development of open source farm management and data tools that enable farmers to control their own data and opt in to opportunities for sharing data. FarmOS, for example, is a free and open-source platform for farm management, planning, and record keeping.

These new paradigms may help farmers to not only recognize the value of sharing data, but to directly realize that value. Enhancing the supply chain’s ability to track back to the source will enable a market that values conservation and soil health.

Centralized and Integrated Federal Data Collection

Another emerging approach to data gathering is to integrate data collection for one topic with other data collection activities. Most commonly, survey data are integrated with administrative data associated with farm program implementation. NASS is actively pursuing this approach where appropriate, and the agency has partnered with the FSA to coordinate the collection of selected data associated with crop production.

Fully integrated data collection is being facilitated in some countries—notably the Nordic countries and Canada—by centralizing the authorities for data collection, through legislation, in a single, national-level data collection agency. The data collection institutions in these countries represent a unique opportunity for the USDA to explore alternative data collection and management models.

Statistics Sweden is a good example of this centralized approach. The agency’s primary task is to “supply users and customers with statistics for decision making, debate, and research.” Not only does the agency serve government needs, they also undertake data collection for the private sector and researchers. About half of Statistics Sweden’s funding comes from government appropriations.

Public–Private Data Collection: Remote Sensing Partnerships

Recent private-sector advances in operational remote sensing applications using public satellite remote sensing platforms bode well for new data collection and analysis methods that are less onerous than current methods (such as extensive surveys) and generate better, spatial explicit, and temporally consistent results.

Mapping Crop Residue and Tillage Practices

Satellite remote sensing is a valuable technology for mapping, assessing, and monitoring the impact of agricultural management practices on soil and water resources over a variety of scales. Remote sensing data have been used to accurately map tillage practices, for example. Van Deventer et al. (1997) achieved high accuracy (93 percent) when mapping tillage practices in a region of soybean–corn rotation in Ohio using Thematic Mapper (TM) shortwave infrared bands.
Another approach to mapping tillage practices focuses on direct estimation of the amount of crop residues in a field using a TM-based Cellulose Absorption Index to infer tillage practices (Daughtry 2006). Reflectance in the shortwave portion of the spectrum (between 1,600 nanometers and 2,100 nanometers) is sensitive to changes in water content, cellulose, and lignin, which have been shown to be related to crop residue cover (Daughtry 2006). Using Landsat 5 data, Sullivan et al. (2008) compared the effectiveness of several crop residue cover indices for mapping conservation tillage practices in a watershed in Georgia. Their logistical regression approach produced accuracies as high as 78 percent. South et al. (2004) mapped no-till practices using Landsat TM data for a region in Michigan, Indiana, and Illinois with a cosine of spectral angle mapping technique. By validating with an intensive transect dataset, they showed that conservation till mapping accuracy can be as high as 95 percent. But they also concluded that the time of Landsat image acquisition, limited by a 16-day repeat overpass, is critical because no-till practices are difficult to differentiate when fields are covered with more than 30 percent crop foliage.

In the past 10 years, data collected by Landsat 7 Enhanced Thematic Mapper Plus technology coupled with logistic regression techniques have been very successful in mapping no-till practices with a high degree of accuracy (>95 percent) for a site in Montana dominated by dryland wheat.

These studies illustrate that satellite remote sensing can be used to map crop residue and tillage systems. However, each study is calibrated for local and atmospheric conditions during the acquisition of the satellite images, hence this methodology is not well suited for operational, large-area mapping of tillage practices. With support from the National Aeronautics and Space Administration (NASA) and the USDA, an operational system called OpTIS for mapping tillage practices was developed. In addition to wide-area mapping of tillage practices, the system uses multi-temporal and multi-year satellite imagery to enable the direct mapping of continuous no-till versus rotational no-till. A public–private partnership led by the CTIC and Applied GeoSolutions used the OpTIS platform to map crop residue, tillage practices, and cover crops annually from 2005 through 2016 for all of Indiana. The data are freely available to the public. This advance in the use of operational remote sensing for wall-to-wall mapping of tillage practices retrospectively and consistently over time is an innovative and cost-effective way to collect agricultural conservation data.

**Mapping Cover Cropping**

While information about the rates of cover crop adoption can be estimated from seed sales, equipment purchases, and program participation, true statistically valid estimates of cover crop adoption and quality are not readily accessible. However, newer remote sensing platforms can easily and cost-effectively track how well and how many farmers are using winter cover crops to protect their soils.

For more than 40 years, data from satellites have been used to map vegetation characteristics across large areas, often with a focus on agriculture. Collected by space-borne optical sensors, reflectance measured in different portions of the electromagnetic spectrum can be used to calculate difference ratios, such as the Normalized Difference Vegetation Index (NDVI), that are then linked with biophysical characteristics, such as vegetation greenness and productivity.

More recently, the relationship between optical reflectance and the characteristics of winter cover crops has been demonstrated (Hively et al. 2009; Prabhakara et al. 2015). These studies have confirmed the relationship between the NDVI and fractional cover, particularly before senescence and winter kill affect the canopy. Biomass can also be reliably estimated from the NDVI. Regional approaches for mapping cover cropping and characterizing the performance of cover crops have been developed. For example, Seifert et al. (2018) used a random forest decision tree classification routine to map cover crops in the upper Midwest for 2008 through 2016.
A recent public–private partnership between the CTIC, The Nature Conservancy, Applied GeoSolutions, and others is using the OpTIS platform to map annual cover cropping and tillage systems from 2005 through the 2017/18 cover crop season for the entire Land Resource Region M (corn belt). The OpTIS products will be used with the DNDC\textsuperscript{17} soil biogeochemical model to estimate the impacts of conservation practices on soil health across the corn belt. The results of this work will be made available to the public at aggregate scale (by watershed and crop reporting district) at https://www.ctic.org/OpTIS.

**Recommendations**

In the 1860s, the primary impetus for data collection at the USDA was to reduce the asymmetry in commodity market information that left farmers, ranchers, and others at a disadvantage when selling or buying commodities. While the USDA has been the primary source for agricultural data for more than a century and a half, information asymmetries have now returned, with the private sector using remote sensing and precision agriculture to collect and analyze enormous amounts of production-related data as a business line. These data are not only being used to improve profitability, they are also used in the micro-marketing of agricultural inputs. The disparity in level of detail between USDA farm production data and private-sector production data is widening.

If one assumes that big data associated with farm production practices has the potential to help improve USDA program development and implementation, then gaining access to the data being aggregated by a few private-sector agricultural firms becomes critical. An excellent example of how private-sector big data could be used to improve USDA programs is the relationship between crop insurance premium rates and conservation practices. Private-sector data could prove to be invaluable in documenting how certain conservation practices can reduce yield variability, which could potentially lead to reductions in crop insurance premium rates for those farmers who adopt these conservation practices.

The following recommendations for USDA-led innovation in the collection, analysis, and dissemination of agricultural data focus on leveraging the USDA’s long history of standard, extensive data collection and encourage the USDA to leverage private-sector innovations in new data collection platforms, and explore using new sources of high-spatial-resolution digital data in the development of new tools for risk management and valuing conservation agriculture.

1. **Ensure Data Privacy**

The USDA already has well-developed systems and policies for protecting the privacy of producer data, because farm-level data play a significant role in shaping USDA program design and delivery. To keep pace with rapid technological developments, however, it is critical that the U.S. Secretary of Agriculture direct agency leaders to collaborate on developing and implementing innovative policies and procedures that enhance protection of private data (as referenced in the 2018 Farm Bill). The highest level of data protection will be critical to establishing data-sharing partnerships with large agricultural data aggregators in the private sector. The Census Bureau’s Federal Statistical Research Data Centers offer an example of highly secure data facilities that enable important research and analysis—the type of research and analysis cannot be done within the USDA’s existing structure.

In addition, the Secretary should contract with a private firm to conduct an annual assessment of data privacy and security at the USDA. The assessment should conclude with recommendations for improving the USDA’s data privacy and security.

The Secretary should also expand the Department’s outreach to groups representing the interests of farmers in their interactions with big data firms in the agricultural sector. The purpose of the outreach would be to gain insights into data ownership issues and to influence data ownership policies at the USDA and in the private sector.
2. Identify USDA Data and Analytical Needs

As part of USDA’s Information Technology modernization initiative announced in December 2017, the Secretary should charge an interagency task force with undertaking a comprehensive assessment of current data collection, analysis, and management by USDA agencies, as well as an assessment of data collection, analysis, and management approaches that are needed to better serve the USDA’s Mission Areas. The task force should identify organizational, technical, staffing, funding, policy, and legislative barriers to improving USDA data collection, analysis, and management. The assessment should also address, at a minimum: data collection partnerships with the private sector; methods to address declining survey response rates; ways to improve interagency coordination related to data collection at the USDA; steps needed to create a USDA data warehouse; and strategies for funding the development of innovative data collection, analysis, and access initiatives. The task force should prepare a report to the Secretary identifying data gaps, with recommendations for improving and modernizing data collection, analysis, and management. An important purpose of this assessment would be to identify private-sector methods and sources of data that could fill identified data needs as well as barriers to integrating private data with USDA data. A requirement for periodic data assessment should also be added to Departmental policy.

3. Develop Data Partnerships with the Private Sector

The Secretary should charge an interagency team with developing recommendations on how to engage in data sharing with the private sector. As part of developing the recommendations, the team should convene a workshop with key individuals from government, the private sector, and universities. The primary purpose of the workshop would be to identify barriers to data sharing, ways to overcome these barriers, and incentives for the private sector to share data while avoiding violating privacy or data-sharing agreements. The Secretary should explore using the authorities in the Federal Advisory Committee Act as well as the existing Federal Geospatial Data Committee to help in developing private-sector partnerships. The Secretary should also explore how to build off of and enhance public–private partnerships, similar to those mentioned in the crop residue and cover cropping example above, regarding the use of remote sensing for enhanced collection of agricultural data. This exploration should leverage NASA activities focusing on partnerships to enhance the use of satellite data in agriculture (e.g., NASA Harvest).

4. Address Declining Survey Response Rates

The Secretary should direct NASS to develop a report with recommendations regarding innovative ways to help ensure the highest possible survey response rates and ways remote sensing data can enhance the use of surveys. In addition, the Secretary should designate a portion of USDA research grant funds to be used for competitive grants related to innovations in statistical analysis.

5. Improve Interagency Coordination of USDA Data Collection

The Secretary should direct the Department’s Mission Areas and agencies to expand the use of USDA Science Council committees to accelerate the coordination of data collection and the adoption of new data collection technology, statistics, and data analysis methods, such as machine learning.

The USDA Science Council advises the Secretary and Chief Scientist on issues related to the USDA’s science agenda. The Council’s areas of interest include food, agriculture, natural resources, energy, and communities. Science Council members represent all USDA Missions, agencies, and program office heads or their senior staff delegates. The Science Council has six standing committees, of which three are relevant to data
collection and analysis: the Emerging Technologies Team, the Remote Sensing Coordination Committee, and the Statistics Committee.

The Emerging Technologies Team coordinates across the USDA to facilitate and accelerate the exploration, adaptation, and adoption of new and emerging technologies for advanced assessments of agricultural production. The team assesses:

- Sensors from space, air, and ground, as well as other sources of data and information;
- Research and development needs; and
- Options for the adoption and integration of new technologies, data, and information into systems used by USDA stakeholders, while ensuring the confidentiality of data collected by the USDA.

The USDA Remote Sensing Coordination Committee fosters closer collaboration across agencies on remote sensing efforts and assures that efforts and investments are not being duplicated, strengths are leveraged, and the USDA is pursuing all reasonable efforts in innovation in remote sensing for agriculture, natural resources, forestry, and food security.

The USDA Statistics Committee facilitates collaboration across agencies on statistical efforts and assures consistency in statistical data on similar variables collected across USDA agencies.

These three committees provide the opportunity to leverage an existing Department-wide platform to help identify and accelerate the adoption of emerging data collection and analysis technology.

6. Create a Data Warehouse

Building on the 2018 Farm Bill, the Secretary should direct an interagency team to develop a data warehouse that will make detailed databases available to university researchers for the analysis of conservation practices and the effect of such practices on farm and ranch profitability, crop yields, soil health, and other risk- and resilience-related issues, all while protecting confidential field- and farm-level data collected by the USDA.

The Forest Service’s Forest Inventory and Analysis program and the Census Bureau’s Federal Statistical Research Data Centers both have data warehouses that could serve as models. These warehouses have resulted in valuable practical research, while upholding the highest degree of privacy protection. The goal would be to allow researchers to have access to detailed agricultural data while preserving the privacy of private landowners.

7. Fund Innovative Data Collection, Analysis, and Access Initiatives

The Secretary should work with the Office of Management and Budget and Congress to obtain funding for initiatives that support enhanced data collection, analysis, and access. Using information from the USDA data needs assessment and other sources, the USDA should develop a budget initiative that supports improving survey response rates, enhancing data privacy and security, developing a data warehouse, and expanding data sharing with the private sector.

Reference List


Endnotes


3. The USDA collects a great deal of other data associated with forestlands, food safety, commodity marketing, trade, grain inspection, nutrition, and other USDA functions.

4. For a detailed discussion of NASS’s data collection and management activities, see Schnepf 2017.

5. The NRCS contracts with NASS to administer the Conservation Effects Assessment Project’s National Cropland Assessment surveys.

6. The USGS, for example, is responsible for collecting, processing, and distributing elevation data for the U.S.


17. DNDC, or DeNitrification–DeComposition, is a computer simulation model of carbon and nitrogen biogeochemistry in agroecosystems. The model can be used for predicting crop growth, soil temperature and moisture regimes, soil carbon dynamics, nitrogen leaching, and emissions of trace gases, including nitrous oxide, nitric oxide, dinitrogen, ammonia, methane, and carbon dioxide.


About AGree

AGree drives positive change in the food and agriculture system by connecting and challenging leaders from diverse communities to catalyze action and elevate food and agriculture as a national priority. AGree recognizes the interconnected nature of food and agriculture systems globally and seeks to break down barriers and work across issue areas. Through collaboration and frank discussion, AGree continues to encourage a broad coalition of interests to build trust, find common ground, and develop shared strategies for achieving transformative change.

Contact us:
1800 M Street, NW, Suite 400N
Washington, DC 20036
(202) 354-6440