AGree
Transforming Food & Ag Policy

USDA and Agriculture Data: Improving Productivity while Protecting Privacy

By Todd J. Janzen* and Laurie Ristino**

July 2018

About the Authors

*Todd J. Janzen is an attorney and founder of Janzen Agricultural Law LLC, a law firm dedicated to serving the needs of farmers, ag technology providers, and agribusinesses. Todd serves as the Administrator for the Ag Data Transparent project, a national effort to bring transparency to contracts between farmers and technology providers. Todd is the former chair of the American Bar Association's Agricultural Management Committee and the Indiana State Bar Association's Ag Law Section. Todd regularly publishes articles on agriculture and technology on the Janzen Ag Law Blog, which are reprinted in Farm Journal, Precision Ag, Progressive Dairyman, and other farm publications. Todd grew up on a grain and livestock farm in Kansas.

**Laurie Ristino is the inaugural Director of the Center for Agriculture and Food Systems at Vermont Law School and an associate professor of law. Previously, Laurie served as a senior counsel advising both the Natural Resources Conservation Service and the Forest Service, as well as in various leadership positions at the U.S. Department of Agriculture in Washington, DC. She specializes in the conservation title of the farm bill, land and natural resource conservation, food security, and ecosystem services. Laurie hold a B.A. from the University of Michigan, J.D. from the University of Iowa, and MPA in public management from George Mason University.

Foreword

AGree has been working collaboratively with the diverse interests that comprise the Conservation and Crop Insurance Task Force for over three years to devise practical strategies for empowering farmers to adopt conservation practices on working lands while maintaining a viable federal crop insurance program. Given the importance of both conservation programs and crop insurance in the 2018 Farm Bill, this paper raises timely questions about how agricultural data can be leveraged to improve environmental outcomes on working lands, maintain the defensibility of the federal crop insurance program, and protect producer privacy.

This paper is authored by Todd J. Janzen, President of Janzen Agricultural Law LLC in Indiana, and Laurie Ristino, Director of the Center for Agriculture and Food Systems and Associate Professor of Law at Vermont Law School. Together, the authors explore ways that agricultural data is currently utilized and explain how a data warehouse at the U.S. Department of Agriculture (USDA) could improve intra-agency efficiency and enable cutting-edge agricultural research.

Safely harnessing agriculture data is critical to improving farm productivity and ensuring a stable future food supply. It can reduce producer risk by providing accurate information about the impacts of farming practices, improve conservation and inputs targeting, and strengthen agriculture policymaking. However, there are still questions and concerns in farm country about how to leverage agriculture data while safeguarding producer privacy. How can USDA harness the power of agriculture data safely and securely to benefit farmers, researchers, policymakers, and tax payers?

The paper explores the potential of agriculture data and offers insights about how new circumstances may enable USDA to strengthen its agriculture data management. While the concepts discussed in this paper will enrich AGree's discussions, they do not represent official AGree positions or the opinions of AGree's Co-Chairs, Advisors, or partners.

We hope you find this paper to be a useful resource.

Deborah Atwood
Executive Director, AGree

Contents

Executive Summary

The last five years have witnessed an explosion of online cloud-based platforms that promise to revolutionize farming. These platforms collect agriculture data, such as yield data, soil type, rainfall, fertility and dozens of other data points across entire fields. The promise from these platforms is enormous, causing agricultural companies like John Deere and DowDuPont to invest millions to create data warehouses. Likewise, the promise has lured new ag tech startups into the data warehouse market from Silicon Valley, to the Great Plains, to the Corn Belt. These cloud-based ag data warehouses all share one feature—they all promise to help farmers produce more with fewer inputs.

One of the ways ag data platforms promise improvement on the farm is through benchmarking field performance with other, similar fields. Because many farmers are leery of sharing ag data information with unknown third parties or even known neighbors, private industry has developed ways to "anonymize" data so that information can be shared with little risk to the farmer. The industry has also stepped in and created an "Ag Data Transparent" seal to recognize honesty and transparency. The result is that many farmers have embraced ag data sharing platforms.

In many ways, USDA has been left out of this on-farm revolution. USDA lacks a common warehouse for the data it already collects, and lacks an infrastructure plan for further data collection across all agencies of the Department. This is detrimental to taxpayers, to agencies within USDA that rely on good data when making decisions, and to universities and researchers that could safely utilize this data to improve farming. This paper examines the USDA's need for a centralized data warehouse and a coordinated data collection strategy that would increase intra-agency efficiency and allow researchers and universities access to study production and conservation data. Of course, the ultimate beneficiary is the U.S. farmer.

What's at Stake?

Agricultural data is the linchpin to improving agricultural productivity, ensuring wise use of taxpayer dollars, and providing for a future food supply. Unfortunately, the lack of comprehensive, useable, and accurate agricultural data is undercutting the ability of policymakers to craft effective agriculture policy.

For example, taxpayers pay 62% on average of federal crop insurance premiums, the main "farm safety net." Like any insurance product, federal crop insurance premiums should be tied to risk. In other words, riskier production practices should result in a higher insurance premium. Higher premiums, in turn, should dis-incentivize producers from engaging in these planting practices. However, premiums are not currently calculated in a manner that fully assesses planting risk. This type of risk assessment would require data sets containing information regarding on-the-ground conditions, such as soil type, hydrology, and conservation practices. Although some of this data exists (such as soil type), other data that would help the federal crop insurance program better predict planting risk, does not.

Meanwhile, traditional agribusiness companies, like John Deere and Monsanto, have long recognized the potential of site specific data as the foundation of precision agriculture to drive more effective agriculture production. These companies are making it easy for farmers to upload data to their cloud-based platforms so that the data can be used for other purposes. For example, in 2013, Monsanto purchased Climate Corporation to help build its data and services business lines. The Climate Corporation uses big data to predict weather risk and then sells insurance policies to producers to manage cropping risks associated with weather.

At the same time, producers have legitimate concerns regarding protecting the privacy of their personal and proprietary business information—when their data is collected by business or government. A 2016 study by the Farm Bureau found that 77% of farmers were concerned about which entities had access to their data and whether it could be used for regulatory purposes.

The Agricultural Data Act of 2018 (the "Ag Data Act"), currently in the Senate-passed version of the 2018 farm bill, includes language to both address the serious data and knowledge gap in federal policy while protecting producer privacy. In particular, the proposed legislation addresses producer confidentiality concerns by incorporating robust legal privacy protections. In addition, the proposed legislation requires any published research to show only aggregate data. Moreover, the legislation clearly limits data use to bona fide research, in part to address producer concern that information is not shared with federal regulators for enforcement purposes.

In the following section, the paper sets forth examples of existing USDA data collection and analysis programs, reflecting USDA's successful track record of protecting producer privacy. In the third section, the paper details the multiple benefits of harnessing the power of big data. In the last section, the paper provides a brief summary of conclusions.

USDA Precedent and Technology to Protect Privacy

USDA has a successful track record of collecting and analyzing landowner and producer data while protecting privacy. Moreover, additional technologies have been developed to make anonymous (or anonymize) individual data to protect identity and privacy. This section provides examples of current USDA data warehouses and also discusses technology to anonymize data that is currently in use to protect privacy while unlocking the potential of big data.

A proven track record of protecting producer and landowner privacy

- **Forest Service Forest Inventory and Analysis Program (FIA):** This program is a census of all forest lands in the United States providing important data to assess the needs of American forests and their owners. The database is used by policymakers, scientists as well as private forest landowners, who own the majority of

American forests. Privacy of landowner data obtained through the annual National Woodland Survey is required by law. The Forest Service only releases data in aggregate to ensure privacy is maintained. See, https://www.fia.fs.fed.us/

- **Agricultural Resource Management Survey (ARMS):** ARMS is the primary source of data for the USDA and the public regarding U.S. agricultural resource, use, cost and farm financial conditions. The survey is conducted by the National Agriculture Statistics Service (NASS) in conjunction with the Economic Research Service (ERS). The survey is a critical resource because it is the only source of data available for objective evaluation of many aspects of U.S. agriculture and rural economies. For example, researchers may download custom reports on farm finances and crop production practices. See, https://www.ers.usda.gov/data-products/arms-farm-financial-and-crop-production-practices/arms-data/

  To ensure confidentiality of producer private information, NASS has strict policies and safeguards in place. For example, NASS processes survey data separate from names and other individual identifiers and allows only authorized agents of NASS access to confidential information. These personnel are subject to fines and other penalties for unauthorized disclosure. See, https://www.nass.usda.gov/About_NASS/Confidentiality_Pledge/index.php

- **Risk Management Agency (RMA):** RMA manages the Federal Crop Insurance Corporation which provides crop insurance to producers. In 2000, the Agriculture Risk Protection Act (ARPA) was passed that required the use of data mining and data warehousing to discover instances of insurance fraud and abuse. To implement ARPA's requirements, RMA contracted with the Center for Agribusiness Excellence (CAE) at Tarleton State. CAE developed technologies to identify anomalous claims using data mining and provides RMA with lists of producers to spot check for insurance fraud. ARPA prohibits the disclosure of information provided by a producer, except in statistical or aggregate form or when a producer specifically consents to such disclosure.

Technology protocols exist that can be employed by USDA to protect personally identifiable information

Tools for anonymization of data used by industry

Many private sector cloud-based platforms allow farmers to view other farmers' data in the form of "anonymized" aggregated datasets. This is accomplished in a way that avoids sharing a farmer's personally identifiable information, or "PII."

There is not currently a universally accepted standard for "de-identifying" agronomic data that removes PII. Still, many private sector ag data platforms have gained widespread acceptance with farmers by removing PII from collected agronomic data to make the remaining data anonymous and shareable among platform subscribers. Similar approaches could be used by USDA to anonymize datasets to make such information accessible to researchers and universities.

As an example, typical data collection from a modern corn planter might include: (a) farmer name/farm name; (b) field name; (c) GIS data (latitude, longitude, elevation); (d) field speed; (e) seed variety; (f) planting depth; (g) population intended and actual; (h) amount of starter fertilizer; and (i) insecticide used. Combining all of these data points across an entire field would yield a unique planting dataset for that particular field. When combined with soil type, yield results, and weather data, the combined data layers could be highly valuable to farmers when making agronomic decisions for the following year (or mid-season to adjust fertilizer or pesticide usage). Of all of these data points, only (a) farmer name/farm name and (b) field name is typically considered PII.

Removing (a) farmer name/farm name and (b) field name from the dataset provides a basic form of de-identification. The remaining dataset would be anonymous and still provide value to researchers who need not know who owns the field or machinery that generated the data.

Because anonymous data with GIS information could still identify landowners by cross-referencing with land ownership databases, private industry has developed other protections that prevent the disclosure of PII to persons running queries on the anonymized datasets. Rather than release the raw data to researchers, many ag data platforms conduct the data queries internally, only releasing the collective, aggregated results. For example, if a researcher sought yield information for a specific corn hybrid in a particular soil type, the dataset holder could query the database on behalf of the researcher and release yield information for that particular corn hybrid and soil type on a county by county basis, rather than provide the raw data to the researcher. Releasing raw data to a researcher is not always necessary to perform the research.

When raw GIS data is necessary to perform research, ag data platforms may also require those querying their aggregated and anonymized datasets to keep any potentially identifying information (such as GIS data) confidential by obtaining consent to share from the farmer and requiring the researcher to abide by the privacy policies established by the data holder.

The Ag Data Act addresses the challenges of keeping data anonymous with specific "Procedure to Protect Integrity and Confidentiality." First, the Act requires any research or analysis published from data warehoused by USDA meet certain privacy and confidentiality requirements. Second, the Act limits the persons who have access to the datasets to persons authorized by the USDA. Finally, the Act contains "protections from release" that would prevent release of data by USDA except as allowed.

The Ag Data Transparent Certification

The ag data industry responded to many of farmers' privacy concerns with ag data sharing with the creation of the *Privacy and Security Principles for Farm Data* (the "Core Principles"). The Core Principles were heralded as an agreement between industry and farm groups, such as Farm Bureau, National Farmer's Union, National Corn Growers,

and American Soybean Association, and National Association of Wheat Growers, and agriculture's technology providers, such as John Deere and The Climate Corporation (Monsanto), to address farmer concerns about expanded use of ag data by technology providers. The Core Principles establish guidelines for collecting, sharing, storing, and transferring ag data.

After establishing the Core Principles, the coalition of farm groups and ag technology providers created the "Ag Data Transparent" seal of approval to recognize technology companies that follow the Core Principles in their agreements with farmers. To become certified, participants must answer a series of ten questions about collection and usage of ag data, then submit those answers to a third party for verification. To date, 13 companies have become certified "Ag Data Transparent."

The Ag Data Transparent certification demonstrates that ag data can be safely shared by farmers with third parties in a manner that respects farmers' privacy concerns while allowing use of the latest ag data technology, including sharing of anonymized data. When implementing the Ag Data Act, USDA can follow many of these same principles and guidelines to similarly assure farmers that ag data will not be misused. For example, collection of data is voluntary, which means it is only done with the consent of the farmer.

The Benefits of Harnessing Big Data

There are numerous public benefits of using big data to improve agriculture policy and programs. This section provides a summary of some of the key public benefits which have informed the drafting of the proposed farm bill legislation.

Producer Risk Reduction

As noted above, big data is being used by private companies to assess production risks and improve productivity at the field level. This analysis is then turned into proprietary products and, in some cases, those products and analytics are sold back to producers. Private companies understand the necessity of big data to manage and mitigate planting risk, especially in light of extreme weather and its devasting consequences on crops. Similarly, the government could better optimize pubic resources to address farming risks if it was able to more fully harness agriculture data. In order to better align the farm bill's safety net with cropping risk—a necessity to ensure a food secure future—it is imperative that government have access to big data and apply the insights gleaned from this information to improve public policy and programs.

Improved Conservation and Resource Inputs Targeting

The dearth of data and analysis impacts the effectiveness of federal agencies like the Natural Resources Conservation Service (NRCS) and their partners when providing conservation technical and financial assistance to producers.

For example, in carrying out farm bill conservation programs, NRCS ranks producer "offers" to install conservation practices in an attempt to fund those offers which will provide the most environmental benefits. However, the actual environmental benefits of conservation practices depend on site-specific variables as applied to broader resources concerns (e.g., reducing agriculture field run-off to address watershed water quality issues). This kind of resource concern targeting requires big data and analytics to achieve the degree of environmental benefits necessary for clean water and healthy soils. Although valuable analysis of conservation practices has been done in some regions

by NRCS through its Conservation Effects Assessment Project (CEAP), a broader research agenda is needed to generate comprehensive insights. Moreover, a lack of access to conservation practice data has hindered public scientific research to the detriment of public policy.

Improved Policy Outcomes

Relatedly, data and analysis are also key to improved policy outcomes. For example, science-based risk assessment at the field level would create a more efficient and fair Federal Crop Insurance Program. As discussed above, farm bill conservation programs could be more effective if big data and analytics were harnessed to improve conservation programs, including culling ineffective practices and informing overall program implementation and amendment through the farm bill. These kind of iterative programmatic improvements are critically needed given the challenges of extreme weather and other factors that are putting unprecedent demands and stressors on our food system.

Taxpayer Benefits

Improving farm bill programs and the safety net to be more effective and efficient provides a better benefit to the public. Indeed, the American taxpayer contributes significantly to farm safety net and cost-share for improving environmental outcomes on private working lands. Consequently, it is important that these policies and programs are continually improved to ensure they are meeting their purpose and providing a good value for the public dollar spent. Given the complex variables involved in agricultural production, the primary way to achieve the insights needed to refine these programs over time is through comprehensive data analysis.

Catching Up with Industry Trends

Significant investment has been made into ag tech platforms in the United States in the last few years. According to AgFunder, investment in "farm tech" reached $2.6 billion in 2017, a 32% increase from 2016.

Improving farm bill programs and the farm safety net so they are more effective and efficient provides a better benefit to the public

Collection of ag data has moved from young startup ventures to maturing platforms. For example, Farmer's Business Network, a startup ag data analytics platform approximately four years old, raised $110 million from investors looking to grow the platform. Granular, a Silicon Valley ag data startup, was acquired by DowDupont for $300 million. The market recognizes the value that ag data analytics can bring to the farmer and more and more farmers are embracing ag data platforms every day.

Value to Public Universities for Conservation Research

The United States' land grant universities have traditionally provided deep wells of knowledge to farmers through extension services. Unfortunately for these universities, today's market is driven by private ag data platforms that collect data and return analytical analysis to farmers outside of public university system. The pace of private industry development of new data analytic tools threatens to leave universities behind, upending their traditional role as the independent researcher for the farmer.

There is little incentive for existing private ag data platforms to integrate and share anonymized datasets with public universities. Without creation of a platform that makes ag data accessible to universities for research, these institutions will fall behind, lose their ability to grow public knowledge and provide cutting edge research for improving farming practices. An accessible data warehouse with USDA could reverse this trend.

Summary

Farmers have been collecting data on their farms for many years, and in recent years they have started to aggregate their information in cloud-based platforms with other farmers in order to harness the power of data analytics. To address farmers concerns with privacy, private industry platforms have developed different methods for anonymizing data.

USDA has also been collecting data for years and has successfully kept this information confidential. However, as farmers are discovering, the real value in capturing data lies in the ability to aggregate and analyze that data.

The time is right to undertake a comprehensive data strategy for USDA that would allow collection of data across USDA agencies to improve programs, sharing of data with researchers to improve our collective knowledge, and providing analytic tools for the farmers to improve their operations. The Ag Data Act provisions in the 2018 Senate-passed farm bill present the perfect opportunity to bring USDA up to speed with the data revolution currently happening on the farm.

Endnotes

1. John Deere's cloud-based data collection tool is the John Deere Operations Center. https://www.deere.com/en/technology-products/precision-ag-technology/data-management/operations-center/ Pioneer Hi-Bred's (now DowDuPont) online warehouse is called Encirca. https://www.pioneer.com/home/site/us/encirca/.

2. Congressional Budget Office, "Options to Reduce the Budgetary Cost of the Federal Crop Insurance Program," p. 5, December 2017.

3. See generally, Woodard, Joshua D. "Soil, Conservation and Crop Insurance." Accessed June 29, 2018: http://foodandagpolicy.org/sites/default/files/AGree_4%20Papers%20Compilation_March2016_FINAL_0.pdf.

4. Upbin, Bruce. "Monsanto Buys Climate Corp for $930 Million." *Forbes*, October 2, 2013. Accessed June 29, 2018. https://www.forbes.com/sites/bruceupbin/2013/10/02/monsanto-buys-climate-corp-for-930-million/#454ec082177a.

5. American Farm Bureau. "Farm Bureau Survey: Farmers Want Control of Their Data." Accessed June 27, 2018: https://www.fb.org/newsroom/farm-bureau-survey-farmers-want-to-control-their-own-data.

6. Agricultural Data Act of 2018, Senate-passed Farm Bill, beginning at section 12054. Accessed June 29, 2018: https://www.agriculture.senate.gov/imo/media/doc/Agriculture%20Improvement%20Act%20of%202018.pdf.

7. 7 USC 1502.

8. This information was compiled by Todd Janzen, the co-author, from interviews with technical officials at a number of existing ag data platforms.

9. American Farm Bureau Federation. "Privacy and Security Principles for Farm Data." Accessed June 22, 2018: https://www.fb.org/issues/technology/data-privacy/privacy-and-security-principles-for-farm-data.

10. Bunge, Jacob. "Agricultural Firms, Farm Groups Strike Deal on Crop Data." *Wall Street Journal,* November 13, 2014. Accessed June 22, 2018: https://www.wsj.com/articles/agricultural-firms-farm-groups-strike-deal-on-crop-data-1415854870.

11. Ag Data Transparency Evaluator Inc. Accessed June 22, 2018: www.agdatatransparent.com.

12. Rissman, Adena et al., "Public access to spatial data on private land conservation." *Ecology and Society*. Vol. 2, No.22, Art. 24 (2017). Accessed: June 29, 2018: https://www.ecologyandsociety.org/vol22/iss2/art24/.

13. Ecklkamp, Margy. "AgFunder: $2.6 Billion Invested in Ag Tech in 2017." *AgWeb,* March 7, 2018. Accessed June 25, 2018: https://www.agweb.com/article/agfunder-26-billion-invested-in-ag-tech-in-2017/.

14. Burwood-Taylor, Louisa. "Farmers Business Network Raises $110m Series D, IPO Likely in 2-5 Years." Accessed June 25, 2018: https://agfundernews.com/farmers-business-network-raises-110m-series-d.html.

About AGree

AGree sseeks to drive 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. The AGree Initiative is housed at Meridian Institute, whose mission is to help people solve problems, make informed decisions, and find solutions to some of society's most complex and controversial issues.

Contact us:

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