

Modernizing Agriculture Data Infrastructure to Improve Economic and Ecological Outcomes

May 2022



Table of Contents

Executive Summary	3
Introduction	4
Current Agriculture Data Landscape.....	5
Challenges for Integrating and Using Agricultural Data	7
Key Considerations for Integrated Agriculture Data Capacity	9
Relevant Models for Sharing and Linking Data.....	11
Next Steps for Integrated Agricultural Data	18

Executive Summary

Modernizing U.S. agricultural data infrastructure will better equip farmers and the U.S. Department of Agriculture (USDA) with tools to adapt, innovate, and ensure a food-secure future given the increasingly dynamic conditions in which the sector operates.

Data innovation is necessary to address a growing number of critical short and long-term food and agricultural issues, including agricultural production, environmental sustainability, nutrition assistance, food waste, supply chain disruptions, and food and farm labor. Though many farmers are already collecting production data about their farms that can help solve these issues, this information remains mostly unavailable to other farmers, policymakers, and USDA due to a number of issues.

This paper focuses on how data innovation can provide farmers and ranchers with better information about their farms; enable research to understand how different farming practices impact profitability, risk, and environmental outcomes; and improve USDA programs to provide better value to taxpayers.

USDA has a vital, yet unrealized, leadership role to play in facilitating data collection, utilization, sharing, and research. The lack of a clear mandate across agencies, some gaps in authorities, and privacy concerns have hindered USDA's innovative use of data, including the department's ability to facilitate needed research to support decision-making. Notable challenges at USDA and for integrating agricultural data include:

- Lack of Consensus, Open Data Standards
- Absence of Consistent System Interoperability
- Misaligned Incentives
- Gaps in Leadership and Governance
- Inconsistent Legal Authority and Interpretation

Given these challenges, this white paper considers key attributes for integrated data infrastructures to improve the current ecosystem for agricultural data sharing. The attributes include farmer and public trust, privacy and confidentiality protections, independence, data acquisition, scalability, stable funding, oversight and accountability, and intergovernmental support. The white paper applies the attributes to explore four

solutions to effective data acquisition, management, and use in other sectors:

- **CENTRALIZED DATA INFRASTRUCTURE OPERATED BY USDA:** A central capacity could consolidate resources and coordination of data standards and systems, but is likely impractical given existing USDA infrastructure limitations, policies, and authorities.
- **CENTRALIZED DATA INFRASTRUCTURE OPERATED BY A NON-GOVERNMENTAL INTERMEDIARY:** A shared infrastructure for managing data operated as a public-private partnership offers the appeal of leveraging government authorities for data protection and resources, and the flexibilities of the private sector, including protecting proprietary information.
- **DATA LINKAGE HUB OPERATED BY A NON-USDA AGENCY IN THE FEDERAL GOVERNMENT:** Ongoing discussions to establish a National Secure Data Service as part of the National Science Foundation offer a model that would present a highly secure environment for linkage, with limits on data use.
- **CONTRACTUAL MODEL WITH RELEVANT PARTNERS:** A contractual arrangement to compile locally collected data, including consolidation with non-governmental or other relevant data assets.

Each of the considered models offers opportunities for collaboration with farmers and other stakeholders to ensure there are clear benefits and to address the shortfalls in the current system. Careful consideration of the trade-offs of each option is critical given the dynamic weather and economic challenges the agriculture sector faces and the potential new economic opportunities that may be unlocked by harnessing the power of data.

Introduction

The existing data infrastructure struggles to help farmers address today's unprecedented challenges, such as extreme weather events and repeated disruptions in global supply chains.

For example, in 2021, fertilizer prices saw a historic rise in prices due to supply chain problems, attributable to the major hurricanes that hit the Gulf Coast, which coupled with the pandemic, shut down the refineries. In 2019, record flooding in the Midwest prevented farmers from planting on 19 million cropland acres. More fundamentally, weather patterns are changing, altering where and how crops can be grown.

Modernization of agricultural data collection, storage, and analysis is key to better equip farmers and the U.S. Department of Agriculture (USDA) with tools to adapt, innovate, and ensure a food-secure future given the increasingly dynamic conditions in which they operate. Data innovation is necessary to provide farmers and ranchers with better information about their farms' productivity and risk; enable researchers to understand how different farming practices affect productivity and environmental outcomes, which can enable ecosystem markets; and drive policy improvements in USDA programs. To realize the strategic benefits of using data, an underlying infrastructure that facilitates the collection, management, sharing, linkage, and protection of data is necessary. While many food and agriculture stakeholders and companies already have modernized data systems in place, others are just beginning to embrace leveraging data as a strategic asset.

USDA has a vital, yet unrealized, leadership role to play in facilitating data collection, utilization, and sharing to meet unmet data needs of farmers and other stakeholders, as well as improve its own program implementation. USDA is pivotal to addressing barriers and challenges to data utilization for several reasons. For one, the department collects enormous amounts of data in its role to support agriculture through both its statistical services and programs. Through a host of farm bill programs, USDA provides farmers with risk management support as well as financial, technical,

and conservation assistance. However, data collected in carrying out these programs are generally siloed within the implementing agency and lacks interoperability with other datasets within USDA. In practical terms, this makes it challenging, if not impossible, to gain insights into how well programs are working in relation to each other and how they can be improved to achieve better outcomes and value for every taxpayer dollar spent.

One prime example is the disconnect between conservation and risk management programs. A growing body of research is finding that implementing conservation practices, like cover crops on cropland, reduces risk.¹ Yet, the Risk Management Agency (RMA) and Natural Resources Conservation Service (NRCS)—the two USDA agencies responsible for crop insurance and conservation programs, respectively—largely implement their programs in isolation from each other. These agencies could improve risk management and conservation outcomes if they worked together to collect, analyze, and apply data insights on the impacts of risk on conservation practices. Improving program performance would also provide a better value for the taxpayer investment, including reducing cost.

Research is another area where modernizing data collection, storage, and utilization would yield much-needed insights. Supporting research on the connection between conservation practices and environmental outcomes (including soil carbon sequestration) and agricultural risk and productivity is crucial to advancing climate-smart agriculture. As noted, there is evidence that conservation practices are effective in improving soil health and farm resilience to climate change and extreme weather events, however, more research is needed to understand how specific practices implemented in different production systems and regions affect carbon

¹See, for example, Bowles, T.M. et al. 2020. Long-Term Evidence Shows that Crop Rotation Diversification Increases Agricultural Resilience to Adverse Growing Conditions in North America. *One Earth*. 2 (3) 284-293; Poeplau, C., and Don, A. 2015. Carbon sequestration in agricultural soils via cultivation of cover crops – A meta-analysis. *Agriculture, Ecosystems & Environment* (200) 33-41; and Conservation Technology Information Center. 2020. Report of the 2019-2020 National Cover Crop Survey. 2020. Joint publication of the CTIC, the North

Central Region Sustainable Agriculture Research and Education Program and the American Seed Trade Association.

sequestration, greenhouse gas emissions, water quality, farm profitability, and risk. This level of specificity is critical to improving USDA conservation and risk management programs as well as catalyzing private ecosystem services markets to compensate farmers for the environmental benefits they create. Concerns about producer data privacy and a lack of industry-standard data sharing architecture and protocols hamper data sharing between USDA and researchers.

Modernizing the national data infrastructure for the agricultural sector is the linchpin to provide critical agricultural insights, improve the effectiveness of farm bill programs, and deliver better value for farmers and taxpayers. Harnessing existing data from government,

industry, and individual sources has the potential for farmers to work in a more productive, streamlined manner and economically empower rural America. Realizing those benefits requires policy change and an orientation to using integrated data for analysis while ensuring important privacy and confidentiality protections are provided. Achieving these goals is possible. This white paper provides context and a synopsis of obstacles for using data more effectively to address conservation and climate change efforts. It then provides an overview of key considerations to address identified challenges and weighs relevant models for agriculture that are used in other public sector contexts against those criteria.

Current Agriculture Data Landscape

Private companies have already proven the value of capturing and utilizing agricultural data, but the collection, integration, and use of data by the USDA has not kept pace.

For example, private companies are using weather, soil, and field data to help farmers determine yield-limiting crop conditions and make production decisions. Others are using microbiology and technology to improve resilience of crops and foster the development of carbon markets. And companies that manufacture agricultural machinery are linking the latest tractors and harvesters directly to the cloud, so that real time data from the field can be instantly collected and used later.

Meanwhile, the collection, integration, and use of data by the USDA to help improve on-the-ground outcomes for farmers and program performance has remained stagnant. The USDA—a massive agency with 29 different agencies and staff offices and 100,000 employees²—collects significant amounts of data to address a diverse set of missions. The lack of a clear mandate across agencies, some gaps in authorities, and the sense that privacy will not be protected have hindered USDA's innovative use of data, including the Department's ability to facilitate needed research.

The USDA's struggle with data infrastructure and management has been well documented. In an

assessment of how the federal government is leveraging data for evidence building, a federal advisory committee detailed the problematic data ecosystem based on feedback directly from USDA.³ The report recognizes that data exists at USDA in hundreds of unconnected silos, requiring employees to make manual data calls to gather basic information for analysis. This, in turn, makes the practice of data-driven decision-making extremely difficult.

The Government Accountability Office (GAO) has also addressed problems with USDA data practices, encompassing Department-wide practices as well as issues with farm-level data collection and analysis. GAO issued Priority Open Recommendations to USDA in July 2021, identifying 11 recommendations across five categories, that the agency remains to address.⁴ Broadly, improving farm-level data collection and management could help address recommendations falling into categories "Reducing Improper Payments" and "Improving Oversight of Federal Assistance and Awards" by providing the information needed to address certain recommendations, including revision of processes for determining eligibility, measuring the

2 USDA. About the U.S. Department of Agriculture. Retrieved from <https://www.usda.gov/our-agency/about-usda>.

3 Advisory Committee on Data for Evidence Building (ACDEB). (2021, October 29). Year 1 Report. Retrieved from <https://www.bea.gov/system/files/2021-10/acdeb-year-1-report.pdf>.

4 U.S. Government Accountability Office (GAO). (2021, July 1). Priority Open Recommendations: U.S. Department of Agriculture. Retrieved from <https://www.gao.gov/assets/gao-21-590pr.pdf>

effectiveness of actions, and ensuring submission of single audit reports from recipients.

In September 2021, GAO pointed out that despite the USDA Farm Production and Conservation (FPAC) mission area launching Farmers.gov in 2018 to provide farmers, ranchers, and foresters with online self-service applications and business tools, FPAC and USDA has not implemented seven of the eight information technology workforce planning activities or developed a strategic plan.⁵ In the same report, GAO

recommended that USDA modernize its IT infrastructure to provide better customer service and to fully address the conservation mission area. More recently, in February 2022, GAO released a report on the implementation of the Market Facilitation Program by the Farm Services Administration (FSA).⁶ The report found payments made to farmers, intended to make up losses for farmers resulting from the 2018 and 2019 trade disruptions, were not accurate enough to be useful for FSA eligibility checks due to data collection design and analysis flaws, among other things.

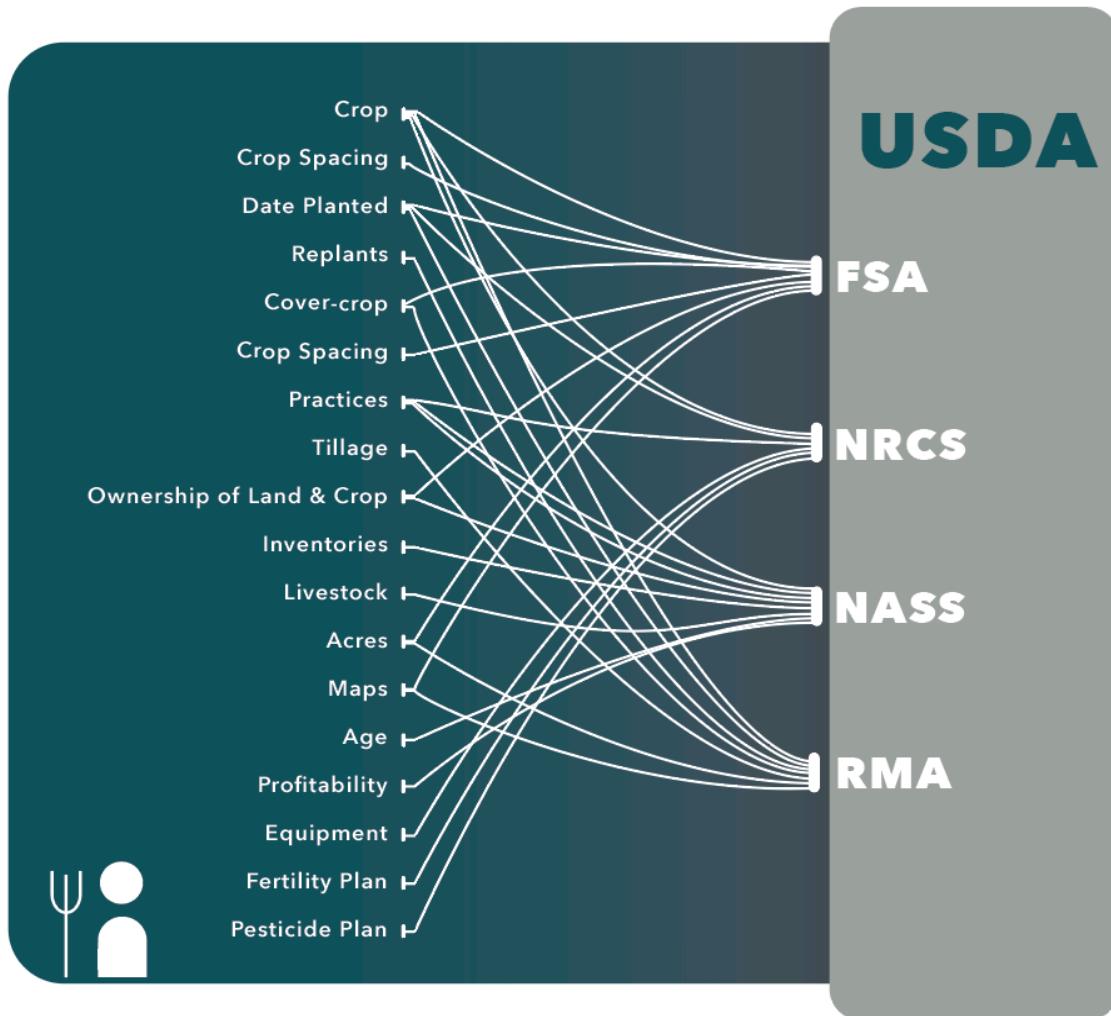


Figure 1. A representation of the data one farmer reports to different USDA agencies, demonstrating that farmers report the same information to multiple USDA agencies. Creating an infrastructure within which USDA agencies could access information reported by farmers via a single platform would reduce the reporting burden farmers face to participate in USDA programs.

5 GAO. (2021, September 23). IT Modernization: USDA Needs to Improve Oversight of Farm Production and Conservation Mission Area. <https://www.gao.gov/products/gao-21-512>

6 GAO. (2022, January 4). USDA Market Facilitation Program: Oversight of Future Supplemental Assistance to Farmers Could Be Improved. <https://www.gao.gov/products/gao-22-104259>

From a producer standpoint, reporting requirements are repetitive. Figure 1 represents the data one farmer reports to different USDA agencies. As depicted, the farmer must report identical data points, such as crops and the date planted, four separate times. Streamlining reporting requirements or creating an infrastructure within which USDA agencies could access information reported by farmers via one platform would remove compliance burdens on farmers and perhaps incentivize more participation in conservation programs.

Despite the well-documented challenges using data at USDA, the department has made some progress in recent years. The 2014 Farm Bill mandated the Acreage Crop Reporting Streamlining Initiative (ACRSI) to reduce the burden of submitting data for farmers and to avoid duplication of data received in different programs across the department.⁷ ACRSI establishes a common framework for farmers to submit acreage reports to USDA, where information is shared electronically and securely between the farmers and relevant mission areas. As part of the initiative, USDA created reporting standards for the framework and published them to industry. ACRSI shows the benefits of data standards to support more efficient, secure, and accurate data sharing across the USDA and its stakeholders.

Since the Foundations for Evidence-Based Policymaking Act of 2018, the department made significant improvements in the practices governing data, establishing a Chief Data Officer (CDO) and assistant Data Officers in each mission area to carry out provisions of the Evidence Act. USDA worked to improve the use of its data, creating the Enterprise Data Analytics Platform and Toolset (EDAPT). EDAPT connects data from 150 sources both outside of and within the department to offer a comprehensive collection of administrative data and a set of standardized, centrally available tools for data analytics. The department-wide dashboard facilitated a shift toward a more data-focused culture and built technical and leadership capacity, inspiring other CDOs to develop similar platforms within their agencies.⁸

With the growing awareness of these data issues and the key role agriculture data can play in improving productivity, risk management, and farmer livelihoods, USDA has been working to modernize and improve its data utilization. However, to truly unlock the potential of data to improve farm productivity and the resilience of rural communities, the department must establish a more effective data infrastructure, which will require addressing gaps in USDA's mandate and authorities across its agencies and programs.

Challenges for Integrating and Using Agricultural Data

The challenges in collecting, integrating, and using data are well known and well documented.

The data available to support policy decisions and practical implementation across the country for advancing farm productivity, addressing agricultural risk, and growing new ecosystem markets are unnecessarily limited. Some of the challenges in sharing, integrating, and ultimately using data are well documented and are not necessarily unique to the agriculture sector. For example, often the limitations are barriers to access from disconnected, disaggregated

systems that are not interoperable.⁹ In other cases, concerns about achieving privacy and data sharing goals at the same time may limit access or data use. Agricultural data also have distinctive barriers and challenges based on the broad range of partners who provide data at the local level and the volume of information collected and shared.

In 2017, the U.S. Commission on Evidence-Based Policymaking surveyed federal agencies about barriers

⁷ USDA Agricultural Marketing Service. (2021, August 5). USDA Market News Reports to Enhance Price Transparency in Cattle Markets. <https://www.ams.usda.gov/press-release/new-usda-market-news-reports-enhance-price-transparency-cattle-markets>

⁸ ACDEB 2021.

⁹ Hart, N., Carmody, K. (2018). Barriers to Using Government Data: Extended Analysis of the U.S. Commission on Evidence-Based Policymaking's Survey of Federal Agencies and Offices. Washington, D.C.: Bipartisan Policy Center. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3927461.

to data access and use. The responses provided from staff in USDA agencies emphasized major barriers around the clarity of existing legal frameworks, availability of funding, and data security and protection.¹⁰ Notable challenges that limit data sharing of agricultural data include:

- **LACK OF CONSENSUS, OPEN DATA STANDARDS.** The adoption of basic, open data standards provides a common framework for data assets to be connected. In the absence of standards, data elements to connect across datasets may be highly error-prone resulting in low match rates and, as a consequence, not useful for decision-making. USDA and other federal agencies have authority to develop data standards, typically through the Chief Information Officer. When needed, the Chief Statistician at the White House Office of Management and Budget is authorized to issue cross-cutting data standards. The lack of consensus and standards is widely recognized as a major barrier for federal agencies in general, and improved processes for identifying where data standards are needed and when to apply standards is highly recommended.¹¹
- **ABSENCE OF CONSISTENT SYSTEM INTEROPERABILITY.** Because state, local, and tribal governments are unique administrative governmental entities, the laws and regulations applied for administrative and operational program activities are diverse and often not designed to plan for interoperability. One respondent from USDA to the Evidence Commission's survey elaborated about the absence of "linking variables" which are used to connect datasets to each other for integration and extended analysis. The linking variables are a basic feature of interoperable systems across public sector organizations, the private sector, researchers, and data owners. Different governmental jurisdictions may also have legal restrictions on what can be shared across programs that further inhibit interoperable systems.
- **MISALIGNED INCENTIVES.** Many efforts to share data require data owners to voluntarily submit without compensation for the value of the data provided or the recognition that data farmers at the most local level bear a cost for data collection and reporting. USDA programs, which provide financial, technical, and other assistance, could be better aligned to link programmatic benefits with data

reporting. Regulatory frameworks can also establish mandatory reporting structures for certain data while establishing a compliance incentive. However, in the agriculture context, this approach is disfavored. Consequently, aligning incentives, program purposes, and data collection goals with the capacity of the data producer or owner is likely critical to addressing data issues.

- **GAPS IN LEADERSHIP AND GOVERNANCE.** Consistent leadership in USDA and other organizations about the role of data and analysis in initiatives ranging from local to national provides a capacity for identifying priorities, resolving conflicts, brokering agreements between data providers and users, and sustaining resources for data initiatives as budgets change. While USDA established the role of the Chief Data Officer and data governance board in 2019 with success in addressing many of these concerns, leadership should be provided by multiple individuals throughout USDA and partner agencies. The great challenge is sustaining initiatives with changes in political and senior agency leaders, which can also be facilitated with clear legal direction and authority about such efforts.
- **INCONSISTENT LEGAL AUTHORITY AND INTERPRETATION.** The statutory limitations in current law that apply to USDA and the ability to create an integrated data infrastructure must be calibrated with what is still possible. For example, the 2008 Farm Bill prohibited USDA from releasing producer information at an identifiable level, but it did not prohibit the collection and analysis of the information at an aggregate level.¹² In practice, confidential research activities that generate summary analysis are still possible, but they are hindered by an overly broad interpretation of the law and lack of an overarching mandate and established processes to support public research. USDA has also previously discussed challenges with interpreting federal law under the Privacy Act about secondary uses of administrative data. The Systems of Record Notices required under the Privacy Act may be ambiguous about whether secondary uses are possible.¹³ Finally, administrators may default in interpretations of existing law to protect and not share data rather than with the reasonable interpretation that data may be shared with appropriate protections. The Foundations for Evidence-Based Policymaking Act that applies to

¹⁰ U.S. Commission on Evidence-Based Policymaking. (2016). Authors' analysis of CEP Public Use Survey File: CEP, Survey of Federal Agencies and Offices [version 1] Washington, D.C.: CEP. Retrieved from: <https://bipartisanpolicy.org/wp-content/uploads/2019/03/CEP-Survey-Analysis-File.xlsx>.

¹¹ ACDEB 2021.

¹² Food, Conservation, and Energy Act of 2008. P.L.110-234. Section 1619. (2008, May 22). Retrieved from <https://www.congress.gov/bill/110th-congress/house-bill/2419/text>.

¹³ Authors' analysis of CEP Public Use Survey File.

USDA provided some new authorities to shift the default to openness and accessibility for research and evaluation activities in the future.¹⁴

- **RESOURCE AVAILABILITY.** Implementing national scale, coordinated, and interoperable data initiatives requires sustained investment in capacity, training, personnel, and financial incentives (e.g., grants, contracts).

Key Considerations for Integrated Agriculture Data Capacity

Integrating data systems and developing shared, secure services for using agricultural data can be achieved using several existing models.

Based upon the specific needs of the agriculture sector, such as focus on advancing farm productivity, addressing agricultural risk, and growing new ecosystem markets, some models may be more appropriate than others.

The overarching goal of modernizing the agriculture data system is to address the identified barriers to data utilization while facilitating efficient, timely, and privacy-protective activities for generating relevant statistics and insights. In view of this goal, there are eight key principles and characteristics to consider when evaluating potential options. The considerations are adapted from principles presented by the U.S. Commission on Evidence-Based Policymaking.¹⁵

- **FARMER TRUST.** To maintain the farmer's trust as the primary data providers, efforts that integrate data need to clearly explain what activities are being undertaken, the benefits of those activities, and the standards employed to maintain farmer privacy. The approach must also include a diverse and representative oversight infrastructure.
- **LEGAL AUTHORITY TO PROTECT PRIVACY AND CONFIDENTIALITY.** For options that rely on the government's infrastructure, comprehensive and clear authority should be provided to protect the privacy and prevent misuse of data that are

accessed. For example, the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) provides statistical agencies in the federal government—including the Economic Research Service and the Agricultural Research Service—data sharing authority and provides mechanisms for both civil and criminal penalties for violations. CIPSEA also directs a series of protections and approaches to implement disclosure avoidance protocols, which reduce the risks of any individual person or entity from being identified in released data files and publications. To carry out the necessary, secure functions, the support staff of an integrated data infrastructure should have authority to access key data sources, technical expertise to clean, curate, and link data, and be able to provide technical assistance to federal, state, and local program agencies and external researchers in using integrated data.

- **INDEPENDENCE.** To help bolster the public's trust in the reliability and accuracy of data, integrated data and research capacity provided in a unit or organization should be to operate apart from policy and related offices in a government agency. This will support objective analyses and enable a stronger culture for protecting privacy and confidentiality. Independence does not mean the unit or organization would operate as a silo or

¹⁴ Foundations for Evidence-Based Policymaking Act of 2018. P.L. 115-435. (2018). Retrieved from <https://www.congress.gov/bill/115th-congress/house-bill/4174>

¹⁵ Hart, N., Potok, N. (2020). Modernizing U.S. Data Infrastructure. Data Foundation.

in the absence of shared priorities; the unit would ensure all its actions facilitate a more open, secure data sharing system.

- **LEGAL AUTHORITY TO ACQUIRE DATA FROM OTHER AGENCIES.** Any entity collecting or combining sensitive and identifiable data should have clear legal authority about the acquisition, management, and use. For example, CIPSEA provides federal statistical agencies with a presumption of accessibility, meaning that administrative records held across the federal government are sharable under this authority unless otherwise legally prohibited. However, the only types of activities that can be conducted with the data under CIPSEA are statistical in nature. One of the relevant features is that information can be shared for generating summary insights but is strictly prohibited in this framework from being shared for regulatory or enforcement actions. Providing clear legal authority reduces operational and administrative challenges in implementing integrated data infrastructures.
- **SCALABLE FUNCTIONALITY.** Scalability has several dimensions and is necessary to accommodate demand for high-quality evidence in the agricultural sector. Integrated data infrastructure needs an IT architecture that can expand in a cost-efficient manner, without significant capital investment. In addition, flexibility for staffing and ensuring personnel have relevant skills; implementation of business practices that are clear and efficient; project approvals for data access that are timely; and processes for cleaning, linking, accessing, and analyzing data that rely on emerging artificial intelligence and machine learning capabilities are key aspects of scaling integrated data capacity
- **STABLE FUNDING.** A secure funding source is needed to facilitate continuity, oversight, ability to meet future demands in an integrated data system. These attributes will also attract and maintain investment from users. Long-term and stable funding could include direct federal appropriations and user fees for those who access and analyze data. Planning at the outset for a clearly articulated and documented business model will help balance the substantial upfront investment and ongoing operational costs.

- **OVERSIGHT AND ACCOUNTABILITY.** For a national integrated data capacity, Congress, the Executive Branch, farmers, and farmers need mechanisms by which they can be assured that data uses are responsible, ethical, and legal. Establishing an independent, expert governing body could provide oversight, but additional parameters for auditing, reporting to Congress, or collaboration with an advisory committee may also be relevant.

- **INTERGOVERNMENTAL SUPPORT.** With interactions across government jurisdictions and existing research support systems, a more comprehensive integrated capacity will need capabilities to coordinate and collaborate across governmental entities beyond the federal government. Such efforts should include a formal mechanism for involving state chief data officers, workforce and employment agencies, and other key data providers and partners.

Collectively these eight attributes represent core capabilities and responsibilities envisioned for an integrated data infrastructure, outlining the role and function it would play along with initial expectations for legal authorizations and policies that may be needed for the entity's success.

Relevant Models for Sharing and Linking Data

While there are different approaches for expanding integrated data capacity, the following section explores the application of several models that have been successful in other policy areas and contemporaneous discussions about data sharing.

Each model holds promise and is relevant for modernizing approaches to analyze agricultural data with leadership provided by the federal government, especially USDA. While the options discussed differ in structure, funding, scope, and purpose, all may be adapted to fit the context of the agriculture sector. Considering the extent of current USDA data collection and the future data collection required to maintain its numerous programs, all models anticipate the department will continue to have a key role in agriculture data modernization.

To help demonstrate the fitness of each model to address the challenges facing USDA, the model will be applied to three specific scenarios described in the above section on the current USDA landscape. Each model below will be applied to (1) the lack of coordination between RMA and NRCS, (2) the replication of reporting as seen in Figure 1, from the perspective of a farmer, and (3) the additional research needed that is critical to better understand conservation practices.

CENTRALIZED DATA INFRASTRUCTURE OPERATED BY THE FEDERAL GOVERNMENT

Historically, the development of national data warehouses served the purposes of government infrastructure well. The federal government provided resources and data standards for implementing many such approaches across state and local governments related to health, human services, and education.

One successful example has been the National Directory of New Hires (NDNH), established under welfare reform in the mid-1990s. The system was established specifically to help identify employment and payment potential for non-custodial parents in the child support system. The system is explicit in federal law and directs states to share data from state systems on unemployment insurance, child support orders, and quarterly wages. The federal government compiles the information, adds data for individuals who gain employment (new hires), then provides capabilities for matching and analysis across state lines.

Funded through a Federal cost-share system to support system operation and maintenance in states, NDNH provides a resource for state-submitted data that are retained in the federal system for up to two years before being removed. Access to NDNH data is

restricted to statutory purposes, but it provides both research and operational uses.

NDNH does have its limitations, including that the statutory data retention period limits longitudinal analysis. The system is also maintained by the agency with a mission for the system's initial purpose, even though the uses of the data expanded substantially over the past 20 years. This means the approving agency is often providing guidance based on its mission rather than with the broader purpose as a service provider for multiple programs.

Other examples of this model include most of the analytical systems at the U.S. Census Bureau (e.g., Longitudinal Household Employment Dynamics) and many reporting programs at the Environmental Protection Agency (e.g., RCRAInfo) where the agencies directly maintain a data warehouse subject to rules in respective authorizing laws.

For the agricultural sector, this model could be designed as such that the USDA operates a data infrastructure that allows for direct submission of data by a range of entities, including USDA agencies, farmers, farmers, and companies/organizations. Following the NDNH model, there would likely need to be clear direction encouraging and incentivizing states to share data from state systems. Clear statutory language could be established to limit the uses of

data, what types of analytical projects could be completed, and the duration of data retention. The centralization of the resources in the federal government may also raise questions about enforcement actions or other potential uses of the data for unauthorized purposes. Mitigating these concerns could suggest that building the capacity under one of the existing USDA statistical agencies with CIPSEA authority would be preferable. This model would also require USDA to implement data standards to existing databases. As discussed above, departmental silos and lack of interoperability—not only between jurisdictions, but across existing data within the department—impedes efficient, data-driven insights across programs. A centralized data repository within USDA would need to address current limitations to USDA data collection as well as collection going forward. This approach could be especially valuable for also enabling improved operational analytics for USDA administrative and management activities.

EXAMPLE 1: RMA and NRCS would store data collected from their respective programs in a central repository operated by the USDA. Upon the development of standardized data collection and reporting requirements, the two agencies would be able to monitor the progress of related programs as well as improve policy outcomes by better coordinating their programs and improve implementation through insights gained through data analysis.

EXAMPLE 2: Individual farmers would need only to submit required data once to state systems regarding agriculture. The state would then be responsible for submitting data from farmers, farmers, and ranchers in their jurisdiction to the central USDA repository. In addition to operating under a federal cost-share system that would ease the financial burden of state data collection, the single submission required from both the individual farmer and the state would further ease compliance burdens.

EXAMPLE 3: The centralized system would streamline and standardize security, privacy, and utilization of farm-level USDA data for public research. Bona fide researchers could incorporate the vast amount of data collected by all USDA agencies to analyze the effectiveness of different conservation practices and policies across regions, among other variables thereby accelerating our knowledge about how well and under what circumstances different types of farming and conservation practices work from productivity, profitability, and risk management perspectives.

CENTRALIZED DATA INFRASTRUCTURE OPERATED BY NON-GOVERNMENTAL INTERMEDIARY

Another approach is to develop a shared infrastructure for managing and warehousing data that is operated as a public-private partnership. These approaches have gained some popularity in recent years because they can achieve the agility of the private sector while retaining the oversight and protections provided by the government.

One such model is the Federal Aviation Administration's (FAA) partnership with MITRE under a contractual arrangement called a Federally Funded Research and Development Center (FFRDC). These approaches that are common for research, defense, and energy topics today, allow for the collection of sensitive information and integrated data while creating a trusted intermediary role between government and the data providers. For the FAA partnership, the FFRDC allows the intermediary to access both government and proprietary data to produce reports and summary statistics relevant for both industry and government.

The FFRDC has been a substantial component of the infrastructure for using integrated data to facilitate safe and effective management of the National Airspace System, among other technical capabilities of the FAA. The approach allows airlines to provide voice recorder and black box data along with reports from pilots into a confidential system.¹⁶ In turn, the FAA can use this information to analyze safety issues for airlines but since the FAA does not have direct access to the airlines' individual records, the FAA cannot use the system or data for direct enforcement actions.

To apply this model, USDA could sponsor a research center that has the capacity to integrate voluntarily reported data from across the agriculture industry as well as all levels of government. The establishment of an FFRDC would provide the capabilities to rapidly launch services and would be competitive under the requirements of the Federal Acquisition Regulations. The competitive component is an important one to both maximize stakeholder input and the efficiency of operational processes, all while benefiting from USDA oversight and partnership.

In practice, the FFRDC could suggest data standards or partner with USDA to issue new consensus data

¹⁶ MITRE. (2009, April). Fusing Aviation Data: A New Approach to Keeping Skies Safer. Retrieved from <https://www.mitre.org/publications/project-stories/fusing-aviation-data-a-new-approach-to-keeping-skies-safer>.

standards when needed across key programs. Additionally, this approach would create a system that facilitates interoperability between existing data collection systems. Although the FFRDC would account for interoperability between jurisdictions and stakeholders, it would need to also account for the current silos and issues with interoperability within USDA. Without addressing USDA's internal barriers to data modernization, an FFRDC would be limited in its ability to analyze programs.

Another approach for implementing an FFRDC could include programmatic incentives for USDA to encourage voluntary reporting to a common platform developed in partnership with the FFRDC. For farmers and farmers (or the companies that collect such data) that may prefer not to share certain types of confidential data with the government, the FFRDC model with a trusted intermediary can offer a helpful solution to facilitate the common interest of farmers, researchers, and government in the analysis of high-quality, timely data.

EXAMPLE 1: Using a USDA-established research center, RMA and NRCS would be able to access data from across the department as well as relevant industry data that may have not been previously reported. Due to the independence of the FFRDC model, reporting entities may no longer have potential hesitation regarding enforcement of the degree to which they are utilizing conservation programs. This would allow RMA and NRCS to develop a broader understanding of their relative programs in practice, draw upon the realities of compliance, and coordinate their programs in a way that may better serve the conservation and risk management missions. These benefits would also be realized if USDA used a common platform developed in partnership with the FFRDC.

EXAMPLE 2: Rather than submit the same data to individual agencies, a farmer could provide relevant information about their business—without fear of sharing proprietary information—to one location via the FFRDC model. Upon development of standard reporting and data collection needs, the individual would need only to adopt the new data collection requirements. These benefits would also be realized if USDA used a common platform developed in partnership with the FFRDC.

EXAMPLE 3: Researchers would benefit similarly from the FFRDC model. With access to USDA data as well as

industry, proprietary, or alternative data assets, researchers could develop a broader understanding of USDA programs in practice and the application of different conservation practices and risk management approaches as applied regionally and by farming type. These benefits would also be realized if USDA used a common platform developed in partnership with the FFRDC.

DATA LINKAGE HUB OPERATED BY THE FEDERAL GOVERNMENT

The Commission on Evidence-Based Policymaking recommended establishing a data linkage hub in the federal government to support research and analytics across a broad range of topics.¹⁷ The Evidence Commission recognized that this infrastructure did not exist in many relevant areas to support policy analysis and research and that a broad infrastructure was needed. In 2021, the Advisory Committee on Data for Evidence Building echoed this recommendation and acknowledged that this model could include multiple data services, including hubs for different policy domains to support an array of stakeholder needs.¹⁸ Though not yet established, this model is envisioned as part of the National Science Foundation, operating under CIPSEA designation, and would help consolidate and combine information securely from decentralized data providers and organizations.¹⁹

While discussions are ongoing about whether the data service would be a FFRDC or a governmental entity, the model is unique relative to other linkage capabilities in government today. Although a data service may be housed in an FFRDC, it is distinct from the FFRDC model discussed above in model #2 in that the centralized hub envisioned by the Evidence Commission served primarily research purposes and may have had more limited value for farmers and farm-level users. The data service is envisioned as a capability that would temporarily link data solely for analytical purposes. Because of the use of the CIPSEA authority and the strong privacy framework it provides, a data service would be obligated to keep confidential any data it uses, and any outputs would be subject to disclosure avoidance protocols that minimize the risk of re-identification for any person or business included in analysis. The combination of the privacy law and the temporary nature of the data linkages makes the data service appealing because of the possibility to bring together data assets in new ways while prioritizing privacy protections.

¹⁷ U.S. Commission on Evidence-Based Policymaking. (2017). *The Promise of Evidence-Based Policymaking: Final Report of the Commission on Evidence-Based Policymaking*. Washington, D.C.: GPO.

¹⁸ ACDEB 2021.

¹⁹ Hart and Potok 2020.

The creation of a data service as recommended by the Evidence Commission would be useful for all sectors, including agriculture, and notably, it could be established administratively today. Serving as a secure, single platform for data linkage, individual farmers, farmers, and agricultural companies could voluntarily contribute to and access the service to improve operational capabilities. Even if the central data service were operated by another federal agency, the USDA could similarly create a support infrastructure that reflects the same principles and practices, especially if there is a high level of demand for services. The data service approach is also one that may be the most likely to incorporate privacy-preserving technologies, such as multiparty computation, in coming years as those technologies become cost-effective and scalable.

While the CIPSEA designation would provide the strong privacy framework mentioned above, CIPSEA is focused on statistical data and perhaps may not be the best fit to serve the purposes of USDA program data for the goals envisioned in this context. That is not to say that CIPSEA is not a strong privacy framework—indeed it is. However, the legal framework also has understandable and known limits on what data can and should be used for that may also suggest areas where the legal framework should not be adopted for data sharing purposes in all contexts. Further potential limitations to this approach are that if the capabilities are authorized under another federal agency, USDA may have less influence to ensure data owners and providers are compensated for efforts to support analysis and that there may be a lack of trust on the part of farmers to engage with a non-USDA agency. Additionally, if a centralized data linkage platform is administered by NSF rather than USDA, there may be significant hurdles to ensure the department complies with necessary standards. The vast amount of data USDA collects and interoperability issues within the agency would require the agency to implement necessary standards to existing databases so they are linkable, as well as to establish operability standards for interagency and intergovernmental data sharing systems. Such an approach would also likely offer limited support for operational and management analysis at USDA given the need to compete for prioritization of analytical activities through a system governed or managed at NSF.

EXAMPLE 1: A data linkage service could provide a single location where RMA and NRCS program data would be collected and accessible for research and

analysis. This would allow RMA and NRCS a broader understanding of their relative programs in practice, draw upon the realities of compliance, and coordinate their programs in a way that may better serve the conservation mission. Unlike the other models described here, the data service would provide temporary linkage and be restricted to the high privacy standards needed for CIPSEA designation. However, CIPSEA is best fit for statistical data, so program analysis that would benefit from administrative data may not be feasible with the model particularly if individual records are needed for further investigation or scrutiny following review of group-level analytics to take further actions for services.

EXAMPLE 2: Operating as a single data hub, an individual producer would again experience simplified reporting compliance by having a new standardized reporting method to submit all relevant program data. The individual would also be confident in the level of privacy protection offered by the linkage model and be assured that submitted data would be unable to be used to enforce program compliance.

EXAMPLE 3: Accessible through a single application process, the data service would provide approved researchers access to all government data across departments as well as other data providers. The wealth of data available to analyze would facilitate new research into various factors impacting conservation practices across the country. Similar to the limitations for USDA's RMA and NRCS, CIPSEA is best fit for statistical data, so program analysis that would benefit from administrative data may not be feasible with the model.

CONTRACTUAL MODEL WITH RELEVANT PARTNERS

One major concern in data sharing models is the degree to which data owners or aggregators are compensated, whether financially or otherwise, for valuable data. The National Vital Statistics System (NVSS) is one such model that calibrates the data compilation progress with compensation for efforts. The National Center for Health Statistics (NCHS) partners with 57 state and local jurisdictions to collect detailed birth and death data from across the country.²⁰ The information is used for rapid health surveillance and for producing detailed statistical analyses of public health topics.

²⁰ N. Potok and N. Hart. (2021). "Practical Steps for Building State Capacity and Infrastructure to Use Data for Evidence-Based Decision Making" in Improving Data Infrastructure to Reduce Firearms Violence. Washington, D.C.: NORC at the University of Chicago, pp. 132-165. Retrieved from

https://www.norc.org/PDFs/A%20Blueprint%20for%20U.S.%20Firearms%20Data%20Infrastructure/Improving%20Data%20Infrastructure%20to%20Reduce%20Firearm%20Violence_Chapter%207.pdf.

In exchange for reporting detailed microdata, NCHS provides funding under contract to compensate for the submission of the data. The contract rates are negotiated by a non-profit consortium for vital records agencies. In addition, NCHS provides extensive technical assistance and support for data providers about the classification procedures, relevant codes and data standards, and processing of the vital records data. In addition to resources provided by NCHS, multiple other federal agencies support the vital records system in exchange for submission of portions of the data files. For example, the Social Security Administration (SSA) receives a subset of the data to compile a simplified "death master file" that is used to match benefits payments to deceased individuals' personal information to reduce improper payments.

Due to the centralized contracting model, efforts are underway to incentivize data providers to quickly convert to electronic reporting and sharing through modernized data systems. In practice, faster data reporting is compensated by federal agencies purchasing the information at a higher rate, yet the model ensures that the data providers are fairly compensated for the data they collect and manage. Once under a contract agreement with the federal agency, data providers would be required to provide data. Data access is also restricted by NCHS, which relies on CIPSEA data protections. Researchers can use a Federal Statistical Research Data Center to access the restricted files. There are also ongoing efforts to establish a virtual data enclave for researchers to securely access restricted data files at NCHS.²¹ This model is one that may be likely to facilitate adoption of privacy-enhancing technologies like multiparty computation, though such approaches are possible under other proposed models as well.

There are limits to the contractual approach, including that it does require some coordination among data farmers to negotiate contract rates. From a federal government perspective, it may also overcompensate the full cost of data collection and management, providing a high premium on data that could be shared through other collaborations. This approach also hinges on the ability of data farmers to not sell confidential or restricted data on an open market, because doing so would limit the ability to conduct aggregate analyses without introducing risks of re-identification or other potential harms that are otherwise mitigated today through standard disclosure avoidance processes and tiered access restrictions.

USDA could apply a contractual arrangement to compile locally collected data, including consolidation with non-governmental or other relevant data assets. USDA could also incentivize farmers to participate in data sharing by more closely linking farm bill programs to the collection of baseline data and providing premiums for more comprehensive data sharing. Under a contractual arrangement, farmers, or consortia, would receive compensation based on the level of data, application of standards, quality of data, and other attributes. The model would in turn incentive data use from federal actors who are paying directly for access and use, which might better encourage frequent summary tabulations and reports for stakeholders. At the same time, the compensation and common data standards encourage interoperability at a local level. However, similar to the other models, existing interoperability problems within the agency must also be addressed to adequately produce cross-program analyses.

EXAMPLE 1: RMA and NRCS could benefit from a contractual model, as they would have new access to compiled standardized microdata that may have otherwise been of lesser quality or granularity. The incentives offered by USDA may also provide RMA and NRCS with a higher quantity of data as well. If USDA tied the agencies' conservation programs with the data collection required through the contractual agreements, RMA and NRCS may be able to manage their programs more accurately and efficiently in an integrated manner to improve risk management and conservation outcomes.

EXAMPLE 2: Compensation for the quantity of data, quality of data, and how quickly reported the farm-level data is to the USDA adds an attractive incentive for individual farmers to not only participate in additional USDA programs but to provide additional data. Specifications of data standards, quality, and level of data established in the contract would also ease the reporting requirements for individual farmers, as they would only have to submit data once and would have guidance on exactly how to do so.

EXAMPLE 3: While researchers would not be directly involved in the contractual relationship with USDA and would have to pay to utilize the data, the data would likely be of high quality and granularity to provide deeper insights into programs across the USDA.

In all four of these models, there are clear opportunities for collaboration with farmers, farm groups, cooperatives, land grant universities, agri-business, and

²¹ Centers for Disease Control. (2021). Data Modernization Initiative Strategic Implementation Plan. Retrieved from:

<https://www.cdc.gov/surveillance/pdfs/FINAL-DMI-Implementation-Strategic-Plan-12-22-21.pdf>.

other relevant entities to ensure the systems have oversight and provide clear benefits to the data providers. Each may also be paired with the development and distribution of training materials responsive to emerging needs to enhance data quality and facilitate access.

Table 1. Options for Integrated Agricultural Data Capacity and Satisfaction of Attributes

Attributes	Option 1: Centralized at USDA	Option 2: Centralized Quasi- Governmental Hub	Option 3: Federal Linkage Hub	Option 4: Decentralized Partnership
Farmer and Public Trust	Traditional USDA mechanisms with established processes to identify projects and benefits	Create new mechanisms guided by federal statistical agency principles and practices	Create new mechanisms guided by federal statistical agency principles and practices	Create new mechanisms guided by federal statistical agency principles and practices
Privacy and Confidentiality	Access and use restricted to statutory purposes and administrative protections, could designate most sensitive components within CIPSEA or use other authority. Privacy Act.	Access and use restricted to statutory purposes and administrative protections, could designate sensitive components as a CIPSEA-eligible agent for additional protections if desired or use other authority. Privacy Act.	CIPSEA-eligible designation and other use limitations. Privacy Act.	Ad hoc data sharing agreements in contracts. Privacy Act.
Independence	Subject to USDA Secretary discretion	Quasi-independent, subject to some USDA Secretary discretion and third-party oversight	Likely quasi-independent if at NSF, subject to some NSF Director discretion and third-party oversight	Subject to some USDA Secretary discretion

Attributes	Option 1: Centralized at USDA	Option 2: Centralized Quasi- Governmental Hub	Option 3: Federal Linkage Hub	Option 4: Decentralized Partnership
Data Acquisition and Management Authority	Limited. Possible administrative expansions with some CIPSEA designations. Computer Matching Agreement Act.	Limited. Possible administrative expansions with some CIPSEA designations. Computer Matching Agreement Act.	Yes, within CIPSEA. Computer Matching Agreement Act.	Limited. Computer Matching Agreement Act.
Scalable Functionality	USDA has existing data linkage capacity, data expertise. Would require investment in IT for infrastructure and ensuring states, localities, farmers have resources to participate	FFRDC partner would be selected based upon expertise, data linkage, and IT capacity. Would require investment in state, local, producer partners to have the data skill to participate	NSF equipped with necessary data expertise and IT to develop processes; if at USDA, would need to create infrastructure. Would require investment in state, local, producer partners to have resources to participate	USDA has existing data linkage capacity, data expertise. Would require investment in state, local, producer partners to have the data skill to participate
Stable Funding	Subject to ongoing appropriations, reimbursable authority, commitments from data providers, and the user community	Subject to ongoing appropriations, reimbursable authority, commitments from data providers, and the user community	Subject to ongoing appropriations, reimbursable authority, commitments from data providers, and the user community	Subject to ongoing appropriations, voluntary participation in contracts, commitments from data providers, and the user community
Oversight and Accountability	Subject to GAO, IG, and congressional committee oversight	Subject to GAO, IG, and congressional committee oversight	Subject to GAO, IG, and congressional committee oversight. May be subject to additional oversight	Subject to GAO, IG, and congressional committee oversight. State and local level oversight bodies
Inter-Governmental Support	Need to strengthen local, state, producer relationships	Flexible collaboration model, need to create new relationships with farmers	Flexible collaboration model, need to create cross-agency relationships	Need to strengthen local, state, farmer relationships

Next Steps for Integrated Agricultural Data

There is much work to be done to improve USDA's data capabilities, but fully-integrated agricultural data is essential to continue U.S.'s place as the world's leader in agricultural production.

While a more robust data infrastructure for agriculture is needed, there are many factors and considerations to weigh in determining how to best proceed. In this white paper, we presented a series of key attributes for weighing four options that each have merits for improving the status quo for the agriculture sector and USDA.

In selecting any of these options for further action, authorizing legislation can further clarify data collection, acquisition, sharing, and protection authorities that are critical for an effective system. In addition, outlining meaningful and realistic mechanisms for oversight and transparency while simultaneously encouraging information sharing about how data are being responsibly used will be important for promoting public trust and accountability.

The USDA has made important strides in improving its data capabilities in recent years, but there remains much room for progress to modernize an infrastructure critical to farmers and policymakers alike. The time has come for USDA and the policy community to consider how to accomplish the joint objective of protecting critical data while also allowing for its use to answer critical questions. Taking action to modernize USDA's data infrastructure will promote much-needed innovation and adaptation by equipping the country's farmers, and policymakers with the information they need about farms, farming practices, and agriculture policy.



This paper was co-authored by the Data Foundation and the AGRee Initiative, with input from an advisory group of experts in agriculture data and data policy.

WHITE PAPER ADVISORS. The following individuals advised on the development of this white paper:

- Joseph Glauber, former USDA Chief Economist
- Todd J. Janzen, Administrator, Ag Data Transparent project
- Amy O'Hara, Director, Georgetown Federal Statistical Research Data Center
- Wade Shen, Chief Program Officer, Actuate

THE AGREE INITIATIVE includes the AGRee Economic and Environmental Risk Coalition (AGRee E2 Coalition) and the AGRee Climate, Food, and Ag Dialogue (CFAD). AGRee focuses on developing innovative and scalable policies and pilot programs that support farmers in improving agronomic and environmental outcomes while adapting to weather variability.

AGRee partners believe there are real opportunities to use federal agricultural policy to incentivize and scale agricultural practices that reduce greenhouse gas emissions, improve soil health, and enhance water quality while reducing farmer costs and improving profitability.

THE DATA FOUNDATION champions the use of open data and evidence-informed public policy to make society better for everyone. Data Foundation is the trusted authority on the use of open, accessible data to fuel a more efficient, effective, and accountable government; spark innovation; and provide insights to the country's most pressing challenges.

As a nonpartisan think tank, Data Foundation conducts research, collaborative thought leadership, and advocacy programs that advance practical policies for the creation and use of accessible, trustworthy data.



AGRee is housed within Meridian Institute, a mission-driven nonprofit consultancy that builds understanding, guides collaboration, and drives action to address our world's complex challenges.

1800 M Street NW, Suite 400N
Washington, DC 20036
202-354-6440
HLair@merid.org