Harnessing Agricultural Data with Standards and Interoperability

Executive Summary

Anthropogenic greenhouse gases (GHG) are the dominant cause of climate change with agricultural activities, contributing to 9%-14% of global GHG emissions[1,2]. Standard land management practices, such as tillage, increase CO_2 emissions while regenerative practices, such as cover cropping, help to sequester organic carbon in soil opening a path to offset GHG emissions. Broadening deployment of these practices can help to reduce GHG emissions associated with crop production and for agricultural soils to serve as a GHG sink via carbon sequestration. However, regenerative practice adoption rates are low; for example, 5.6% of growers in the US planted cover crops, <25% rotated their crop, and <35% practiced no-till from 2017 to 2019, and less than 1% of all US acreage was managed using all three practices[3].

One solution is to leverage carbon markets to provide incentive payments to deploy management strategies that minimize GHG emissions and maximize carbon sequestration and its co-beneficial outcomes to soil and crops. To ensure long-term success at a global scale, carbon markets must use meticulous and transparent protocol standards for GHG mitigation and CO_2 sequestration to generate high-quality carbon credits, and growers should be directly compensated for the carbon credits they generate on their operations from carbon registries. Precise, verifiable, and traceable soil and agronomic data must be collected from grower fields and grower records to keep uncertainty low. Agronomic data collection is a significant requirement, as it includes planting, tillage, inputs, amendments, compost, tiling, irrigation, and harvest data for multiple years for a grower and/or landowner to receive payments. Needed documentation for these data require a significant investment from both the grower and the project developer, both in interview time and in the effort to properly clean and assess the data for quality. Furthermore, the burden of data collection remains a challenge for growers as many farm records remain undigitized.

The United States Department of Agriculture (USDA) is well-positioned to reduce this data burden for growers, researchers, and carbon market stakeholders alike. The USDA has an extensive repository of agricultural data; however, it is often challenging for producers, researchers, and organizations to access and leverage their data to generate practice recommendations or support ecosystem service markets due to fractionated and non-standard land management practice data collection, limited dataset interoperability, and the overall regulatory burdens on Natural Resource Conservation Service (NCRS) staff.

In this brief, we propose the following policy solutions to take advantage of the USDA's long-standing relationships with growers, researchers, and leadership in agricultural research to leverage the vast landscape of agricultural data to support grower incentive payments in carbon markets:

1. Implement universal data standards and a comprehensive data trust model to standardize data collection, share interoperable data between mission areas, and promote a culture of interagency collaboration around data.
2. Promote advanced data collection tool adoption and use, for example, remote sensing technology.
3. Engage farmers and ranchers through novel data-sharing incentives and support ecosystem service market development to add value to their operations.
4. Establish funding for an environmental claims clearinghouse and technical assistance to farmers and ranchers.

These policy recommendations will re-frame agricultural data as a vital public resource that can be leveraged for grower and public needs while maintaining data confidentiality and sovereignty for producers.

[1] Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems, 2019.
[2] United States Environmental Protection Agency 2020 "Sources of Greenhouse Gas Emissions", 2020
[3] Indigo Ag, Regeneration on the American Farm Progress report, 2020.

Introduction

The Biden Administration and Congress have a singular opportunity to advance climate change mitigation and adaptation goals through agriculture while improving economic outcomes for producers by promoting improved soil health as a climate change and natural infrastructure solution. Leveraging the ability of agricultural soils to mitigate GHG emissions and sequester atmospheric carbon is not a novel concept as agricultural soils estimated to have an impact potential of 0.25 Gt/year CO_2 in the United States and 3 Gt/year CO_2 globally, second only to bioenergy-based carbon capture according to the National Academy of Sciences (National Academy of Sciences, 2019). To realize this potential, we must promote agricultural management strategies that minimize GHG emissions and maximizing carbon sequestration as a primary goal of federal agricultural climate policy. This goal can be met through incentive payments for adopting beneficial practices, providing outcome-based payments for soil carbon sequestration and other emissions reductions, and aligning subsidy, crop insurance programs, and lending practices. To be successful, high-quality agricultural and environmental datasets are required to demonstrate the efficacy and associated benefits of key agricultural practices across diverse soil types, cropping histories, and climates. This requires that agricultural and environmental data infrastructure and sharing capacity be modernized on a national scale.

However, we are currently in the 'wild west' of agricultural data. Myriad players in the agricultural data space – farmers and ranchers, federal agencies, research institutions, and private enterprises – produce, collect, and collate discrete and overlapping data types. Data collection tools such as producer surveys and remote sensing technologies are often disconnected, with the agronomic and environmental data being collected, stored, and analyzed by different parties. Ultimately, producers pay the price of these inefficiencies through duplicate data capture between producer programs – both private and public. For example, producers seeking advanced technical support or aiming to enroll in a carbon program must spend hours, or even days, providing information that has often already been collected and stored by federal agencies, conservation districts, or private crop advisors. The federal government can play a pivotal role in minimizing this data burden for producers, improving data quality, and unlocking research and market value across the agricultural sector. A recent example of government initiatives to reduce the burden of Federal reporting on growers is the Acreage Crop Reporting Streamlining Initiative (ACRSI) from the Omnibus Appropriations Bill, which allows farmers, insurance agencies and service providers to communicate reporting data electronically with the USDA.

The USDA is the sole entity with consistent records and access to farmers across the US and has already demonstrated the effect of increased data exchange efficacy in crop insurance. The USDA is best positioned to centralize and lead this effort and concurrently improve data tool accessibility and utility for farmers and ranchers to support the transition to production systems with improved ecological, economic, and socially equitable outcomes. The USDA is already a role model for data sharing having implemented the Agricultural Data Commons which functions as a centralized data registry and repository, providing access to agricultural research data produced from USDA-funded projects.

Here, we outline a series of policy recommendations to facilitate modern data collection that address barriers, ensure interoperability, and provide a foundation for improving environmental and economic outcomes for producers. We believe that these are essential steps to help reduce the data collection burden for US farmers and ranchers, promote cross-functional collaborations between all agricultural players, and help the US achieve its climate change mitigation goals in accordance with the Paris Agreement.

Current Barriers to Agricultural Data Infrastructure Modernization

The following challenges represent the immediate roadblocks to improved data accessibility and use across the stakeholder groups:

Fractionated and Non-Standardized Practice Data Collection**:** Current models of data collection all involve producers or agencies entering data into a system owned and operated by a central agency or service provider. If producers want to access these data, they need to navigate these disparate, and often complex, systems. Practice data collection appears to be unevenly conducted between agencies and/or data collection years. For example, a cover crop planting question added to the Farm Service Agency (FSA) survey in 2015 has not been consistently collected. In some regions, county staff ask the question, while others do not. Whether this is due to differences in understanding of cover crop guidelines, practice familiarity, or training, the result is uneven data collection. Uneven data collection produces lower-quality and non-aggregable datasets because of these 'small' differences in data collection method or structure. Some of these differences may be attributed to disparate data uses, and thus may be unavoidable. However, common use cases exist, and data collection methods and structures can be aligned to support interoperability and reduce the costs and complexity of collaborative projects when organizations, technical service providers, and agencies try to share data.

Dataset interoperability**:** Each USDA agency collects data using different data dictionaries and area units. As a result, datasets across agencies are difficult to combine and harness for analyses. Modernizing the approach by using consistent data dictionaries and area units would facilitate greater agricultural insights, for example elucidating the effects of key GHG mitigating land management practices on farm productivity and profitability. Improving dataset interoperability would also address the issue of multiple agencies asking producers for the same information, reducing the data-sharing burden to farmers and ranchers.

Regulatory Burden on Natural Resource Conservation Service (NRCS) Staff**:** NRCS data collection helps to ensure regulatory compliance with conservation programs including the Environmental Quality Incentives Program (EQIP) and the Conservation Stewardship Program (CSP). These programmatic regulatory requirements place a large burden on NRCS county office staff, requiring them to invest a significant amount of time to meet all legal requirements. This limits the ability of staff to allocate more time for producer outreach and education, for example educating and/or providing technical assistance to producers on conservation practices. Solving these issues would reduce the number of services, implementation costs, and learning curves that are now significant barriers for producers who strive to optimally manage their farms and ranches.

Landscape of Data Types

Several federal agencies have led and maintained strong agricultural databases with data enabling intra- and inter-agency activities and can provide immense value to public private collaborations. Table S1 summarizes the data types and their value by federal agency. Modernizing data collection and analysis across these agencies would enable the USDA to also incorporate advancements in precision agriculture and satellite technology, which has created rich sources of agricultural information.

Some private companies have developed their own precision agriculture databases and are offering technical assistance services to farmers, others are aggregating accessible data to develop ecosystem service products and create carbon sequestration and other ecosystem benefit marketplaces. These efforts have traditionally been led by the USDA, and we believe the USDA can provide enormous value by supporting the standardization, modernization, and democratization of its data platform, while providing protections for those whose data is supported. Today, growers can gain limited access to their data through the FSA Farm+ platform, but modern data

onboarding processes seen in other parts of the agricultural ecosystem are not available. By allowing private companies to connect to the USDA's grower data directly through APIs, the government would advance data interoperability, grower involvement in sustainability programs, and better farming data management. Table S2 summarizes current data collection methods and proposes alternative API-driven data sharing methods. In these scenarios, data would still be controlled by the growers and only shared with third parties with explicit consent.

Many growers today have experience with similar systems. For example, modern farm machinery produces copious valuable agronomic data that is stored in third-party cloud accounts (e.g., MyJohnDeere). It is commonplace for other agricultural technology platforms to allow growers to onboard these data – for example, porting their harvest data into alternate software they use to calculate their overall profitability. Growers can both grant access to their data and control which data the systems can access in an ongoing manner. We urge the USDA to consider opening access to grower data in this way to advance modern farming and carbon programs.

Policy-Focused Recommendations for Agricultural Data

We believe the USDA should play a central role in encouraging data standards and data interoperability to unlock value for farmers, ranchers, and researchers. Key opportunities include harnessing and sharing high-value agricultural data, leveraging and expanding the National Agricultural Library Ag Data Commons, an environmental claims clearinghouse, and underpinning this work with data standards, statement of principles, and a data trust. We propose the following policy recommendations to achieve these goals.

Implement universal data standards and a comprehensive data trust model to standardize data collection, share interoperable data between mission areas, and promote a culture of interagency collaboration around data:

Data trust and standards: Data trust models have been used by federal agencies, such as the Federal Aviation Administration, to collect and store both public and private entity data for analyses that improves decision making across the sector and benefits all parties. Paramount to the success of data trust models is the ability to confidentially store data and ensure the privacy of farmer and rancher personally identifiable information. Creating uniform data standards is of utmost importance to achieve this goal, as is education for producers about agricultural data ownership to facilitate new conservation value for farmers and ranchers. The USDA is uniquely positioned to outline and implement standards for grower and agronomic data.

The USDA already collects and maintains clean, standardized data related to annual crop insurance through both the FSA and RMA offices. This type of scale and ubiquity is rare in agriculture. This data set is particularly useful as a potential standard and is widely used by most row crop farmers in the United States. Additionally, the FSA collects and maintains clean field boundaries from growers each year. This information lives in the common land unit (CLU) database that is used for the purpose of enrolling in federal insurance programs via the approved insurance providers (AIPs). Fields are the container for agronomic data and thus a critical component of interoperability across data sources. Unfortunately, fields are very non-standard today, as they are defined by somewhat arbitrary boundaries and names that can change year-to-year or for different uses. As a result, the basic building block of agronomic data is unlikely to be consistent across systems and it impedes data sharing. Standardizing field boundaries is key to improving data interoperability. Additional improvements would be to host data in machine readable formats accessible through tiered access services for secondary uses according to "FAIR Guiding Principles for scientific data management and stewardship." The digital tools which are required to make agricultural data useful for farmers and ranchers should also be published as open-source code. Efforts to create a digital system should be done in collaboration with existing entities that are supporting similar efforts. Incorporating additional research institutions with open data repositories that interoperate with existing open-

source software, hardware, and remote sensing technologies in a shared data structure with USDA will support producer specific recommendations and facilitate the voluntary contribution of high quality, high resolution data into USDA repositories.

Data sharing: Data sharing can facilitate advanced scientific insights which can inform agricultural decision-making supporting economic and environmental farm resilience. Once data has been collected, they should be entered and stored in systems that can exchange and make use of the same information easily, facilitating systems interoperability. The USDA can leverage and expand the Ag Data Commons to use digital and automated data collection tools in collaboration with external partners. This approach should also enable farmers and ranchers to maintain ownership of and protect their data while supporting the use of their data for policy analyses, research studies, and other initiatives (see the OpenTEAM case study below). We support the USDA's current agriculture agenda which includes deployment of precise, accurate and field-based sensors to collect information in real time to visualize changing conditions and respond automatically with interventions that reduce risk of losses and maximize productivity.

To make meaningful insights of the automated use of sensors, data, and visualizations at scale, we encourage the USDA to continue to support interoperability of public tools with open-source technologies. For example, existing the Ag Data Commons can continue to house and visualize data, and be expanded to house cloud-based computing infrastructure and shared software services to accelerate the rate of soil health discoveries and impact on the ground. Clear data standards and statement of principles would continue to build upon this digital system that allows producers to store their data and auto-fill electronic surveys distributed by USDA agencies, certification programs, ag tech companies, and universities. Such a system would reduce producer burden, streamline data collection, and provide producers with greater ownership and transparency over their data. The system would allow producers to easily track their operational data over time and lead to better operational insights. Further, this system would also increase data accuracy as producers would be incentivized to provide accurate information for agency use.

Promote advanced data collection tool adoption and use:

Technology is a necessary enabler for comprehensive data collection; the USDA should oversee a collaborative stakeholder and interagency effort to ensure the best available technologies are used. These technologies should include farm level electronic recordkeeping, in-field testing, remote sensing, climate and weather datasets, agro-ecosystem models, and site-specific decision support tools. Specific attention should be given to satellite data and remote sensing solutions, which can provide and enable analysis of agricultural data at scale. Producers should be engaged to explore ways in which they can benefit from remote sensing solutions, including harnessing resulting insights and reducing data collection and reporting burdens.

Technologies should maximize the efficiency of producer level data collection by streamlining the flow of data collected across federal conservation programs. While technology such as remote sensing, artificial intelligence, low-cost spectroscopy, drone imagery, internet of things sensors, precision agricultural data and environmental models have all rapidly increased in sophistication, accessibility, and lowered in cost, individual components of these technologies were not designed to function as a coherent ecosystem. Despite many advances there is limited adoption of advanced data collection tools, and what tools have been adopted are often disconnected and siloed by different parties. Though different systems and models require different inputs, it is possible to develop standards and systems that allow data to be entered once by a single party yet harnessed by multiple parties for varied use cases. Inter-agency and public-private standards governance are critical. A planned standards development process should include facilitated exchange during the design process and incorporate the full and evolving technology ecosystem.

To ensure data collection is fair and equitable, the technologies employed should facilitate participation by all producers regardless of operation type or scale, and producers should have a mechanism to influence the use of emerging technologies. New and emerging technologies will ultimately enable the scaling of site-specific decision support tools that identify the best opportunities for reducing emissions, increasing soil carbon sequestration, and monitoring the results of these activities over time.

Engaging farmers and ranchers to add value to their operations through novel data-sharing incentives and support ecosystem service market development:

These data standardization and interoperability goals cannot be accomplished without grower participation as it is critical to engage the core constituency when considering innovations or reforms to agricultural data collection. Many data collection efforts are voluntary and, therefore, require the cooperation of agricultural producers who are asked to report data to federal agencies, commodity organizations, NGOs, universities, and private organizations, sometimes with no clear benefit to their businesses or decision-making processes. This reporting burden has created a culture of skepticism and reluctance; producers may be wary to provide additional information to federal agencies and are often unwilling to make their data widely available for research efforts.

New strategies and incentives for data sharing should be explored to create a compelling value proposition for farmers and ranchers. Agricultural producers could be convened to explore what incentives would be meaningful and encourage them to share their operational data more willingly. This meeting should explore how federal agencies can go beyond using data to satisfy the requirements of technical and financial assistance programs and demonstrate the value of agricultural data to producers and other stakeholder groups. For example, agencies could provide aggregated and synthesized information back to producers to help inform on-farm decision making. This is a critical step forward to ensure equity, inclusion, and data access so that operations of all sizes, types, and land tenure situations can use these data tools, contribute to climate change mitigation and adaptation, and access to emerging incentives. Demonstrating that producers can benefit from their data can catalyze the expansion, integration, use, and analysis of agricultural data.

We encourage the USDA to explore new incentives for data sharing to create a compelling value proposition for farmers and ranchers while protecting their data. Over the next few years, ecosystem services markets (ESM) continue to grow and will remain a major lever through which producers are incentivized to adopt conservation practices that improve soil health, water quality, and sequester atmospheric carbon. We encourage USDA to support a diversity of public and private initiatives to value environmental services. The USDA and an Agricultural Data Commons could support ESM efforts, supply chain claims, and outcome-based incentives into programs like CSP.

Establish funding for an environmental claims clearinghouse and technical assistance to farmers and ranchers:

We encourage the USDA to establish and provide long-term funding for a non-federally governed environmental claims clearinghouse to enable review of stacked environmental claims made by diverse public and private marketplaces and incentive programs. The clearinghouse would clear potentially conflicting contracts and claims across field boundaries and contract terms and should be governed by a diverse board overseeing and assuring the integrity of the service. This is important to assure practice additionality, data integrity and interoperability, and prevent double counting in carbon markets.

Finally, we ask that the USDA prioritize Technical Assistance and Cost-Sharing for Climate Smart Agriculture so that producers can make the most of the robust data that will be available. Technical assistance is fundamentally necessary to enable farms to engage in adaptive management planning focused on soil health. This enables farms to identify the most environmentally and cost-effective practices at the outset, allowing for more efficient adoption of soil health practices over time. Novel program priorities in EQIP and CSP would help farmers to adopt

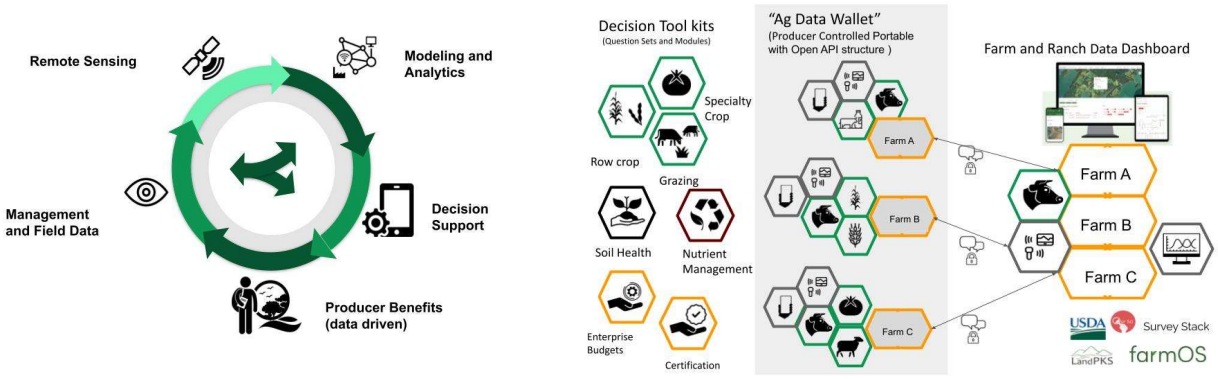
new practices that reduce GHG emission or increase soil carbon sequestration. USDA should increase the funds available for both Conservation Technical Assistance (CTA) and Technical Service Providers (TSPs) to support farmers taking action to reduce GHG emissions, increase soil carbon sequestration, and monitor and report on the results of these activities. Supported activities should include conservation planning and goal setting and increase payments for these activities to a level that creates an adequate incentive for farmers to participate.

Case Study: Open Technology Ecosystem for Agricultural Management (OpenTEAM)

OpenTEAM is an example of a farmer-driven, interoperable approach that provides farmers and ranchers with the best possible agricultural data to improve their soil health by leveraging public and private resources. Their collaborative approach enabled the creation of custom data interfaces and decision support tools that help growers to choose appropriate field methods in alignment with the underlying field data. This reduces sources of error and foments the utility of future observations. OpenTEAM proposes a system in which farms and local trusted agricultural groups host their own farm data and make relevant fields available both to government agencies and technical service providers to enable farm and ranch data portability. These data would be available to other growers on an "opt-in" voluntary basis, and to researchers on a case-by-case basis.

For example, a producer controlled "Data Wallet", would enable individual farm data collected through an ecosystem services payment program to be treated as confidential, and allow the producers to make their data available to researchers through an open API. The data standards and interoperability policy recommendations outlined here would make these data easier to collect in a consistent way to reduce user and data errors. Improved data tools and intake processes can also reduce USDA staff data collection burden and enable greater focus by all stakeholders on generating and sharing high-quality agronomic data, helping to ensure quality and actionable data outputs.

Figure 1. Agricultural Data Wallet and Data Transfer



Conclusion

We strongly believe that these recommendations will have significant immediate and long-standing impact on grower engagement, data sharing, conservation outcomes, and benefit production. These outcomes may also lay essential groundwork to continue to address economic development, economic recovery, technical job training, and rural infrastructure revitalization concerns within the United States. These recommendations will require significant collaboration across the public and private sectors, not only to implement these goals but also to ensure that all stakeholders are committed to continuously contribute to, utilize, and sponsor the effort. Our team of data scientists, modelers, agronomists, and policy experts is at your disposal as you advance these efforts.

Acknowledgements

This piece was developed collaboratively by Woodwell Climate Research Center and Indigo Agriculture. The authors would like to thank all our contributors and reviewers for their time and effort spent contributing to this piece. Your comments and suggestions helped us to refine and improve this work. The authors would like to specifically acknowledge Dorn Cox at OpenTEAM for their review.

Contacts: Woodwell Climate Research Center: David McGlinchey, J.D. (primary) and Jonathan Sanderman, Ph.D. (technical). Indigo Agriculture: Kathryn Meng Elmes, Ph.D. (primary) and Ashok (A.J.) Kumar, Ph.D. (technical)

Appendix

Table S1: Data Overview: USDA, Risk Management Agency (RMA), Farm Service Agency (FSA), National Resource Conservation Agency (NRCS), and National Agricultural Statistics Service (NASS).

Data type	Agency	Added Value Use
Crop Type	RMA, FSA	• Insurance • Finance • Decision tools • Research • Ecosystem Service Markets
Plant and Harvest Dates	RMA, NRCS	
Yield	RMA, NASS	
Field Location and boundaries	RMA, FSA, NASS, NRCS	
Livestock	NASS	• Decision Tools • Research
Economics	NASS	
Conservation Practices	NRCS	• Insurance • Decision tools • Research • Certification • Ecosystem Service Markets
Environmental Assessments	NRCS	• Decision tools • Research • Certification • Ecosystem Service Markets
Conservation Planning	NRCS	

Table S2: Current and Application Programming Interface (API)-Based Alternative Data Collection Methods

Agency	Current Data Collection Method	Alternative Collection Method
Farm Service Agency (FSA)	Producer-led data collection in person at the local county office in person to sign and certify the accuracy of their reports. Electronic reporting recently incorporated including yield and acreage data from precision agriculture equipment.	OpenAPI: collected farmer controlled portable data server to match field boundaries with other agencies, and software uses.
National Agricultural Statistics Service (NASS)	Paper surveys are mailed to producers which are returned via mail, email, or electronic portal. Surveys may be annual, quarterly, and/or survey only a subset of producers. If producers delay submitting responses, they are contacted via phone or in person visit for follow up.	OpenAPI: collected farmer controlled portable data server to auto populate survey forms. Follow up with paper surveys and calls
Natural Resource Conservation Service (NRCS)	Conservation programs use two different tools to process landowner applications and assess and prioritize conservation programs. NRCS began testing the Conservation Assessment Ranking Tool (CART) in 2020	OpenAPI - collected farmer controlled portable data server connected with CART and Conservation Desktop
Risk Management Agency (RMA)	Correspondence between a producer and their crop insurance agent in person, over the phone, or via email.	OpenAPI: collected farmer controlled portable data server auto populated with gaps filled in with agent in person, over phone or email.