Office of the Chief Information Officer
U. S. DEPARTMENT OF AGRICULTURE

DATA STRATEGY

US Department of Agriculture
Fiscal Year 2021-2023

Table of Contents

Executive Summary 3

USDA Leadership on Data and Analytics 4

Achieving USDA's Vision 5

Strategic Goals 5

Goal 1-Data Governance and Leadership 6

Data Governance and Data Leadership

Assistant Chief Data Officers

Data Stewardship

Dashboard Governance

Goal 2-Data and Analytics Workforce 8

Centralized Analytics Teams

Workforce Skills Assessment and Development

Fostering Communities of Sharing and Education

Goal 3-Common Data and Analytics Platform 10

Data Warehouse

Enterprise Analytics Tools

Data Catalog

Geospatial Data

Streamlining Data Collection Methods and Common Customer Records

Goal 4-Open Data 13

Open Data Strategy

Conclusion 15

The US Department of Agriculture touches each American citizen every day through its work with America's farmers, ranchers, national forest users, rural communities, consumers, trade partners, agricultural industry, and the public. On behalf of these customers and the American citizen, USDA aspires to become the best managed agency in the Federal government by focusing on being facts-based, data-driven, and putting customers at the center of all that the Department does.

USDA's ability to leverage its data as a strategic asset and to take a leadership role in the changing agricultural data landscape are key success factors in the Department achieving this vision. Data has been at the heart of its mission since 1861, when President Abraham Lincoln proposed in his first address to Congress the establishment of an "agricultural statistical bureau" and the Department's first annual report included as its first objective the "collecting, arranging, and publishing of statistical and other useful agricultural information." Today, within its more than two dozen agencies and departmental functions, USDA generates and maintains many petabytes of data across hundreds of applications and data stores. USDA has an opportunity to harness its vast data assets strategically to improve internal decision-making and efficient use of resources, maximize the impact of citizen-facing programs, and provide the public and private industry with easy access to data that can solve national problems and drive innovation.

USDA is committed to achieving this goal in alignment with Federal-wide initiatives such as the Federal Data Strategy and the Foundations for Evidence-Based Policymaking Act. This guidance requires that:

- Open government data assets be machine-readable
- Agencies develop and maintain a comprehensive data inventory for all data assets created by or collected by the agency
- Agencies designate a Chief Data Officer (CDO), who shall be responsible for lifecycle data management and for ensuring that the agency maximizes the use of data for the production of evidence, cybersecurity, and the improvement of agency operations.

USDA has made significant strides over the past three years to foster greater data-driven decisionmaking across the Department.

> Looking ahead to FY2021 through FY2023, USDA will build on this progress to remain at the forefront of data and analytics innovation. This progress will be driven by goals focused on improving USDA's data governance and leadership, data and analytics workforce skills, data and analytics infrastructure and tools, and its ability to share data with appropriate stakeholders and with the public.

USDA Leadership on Data and Analytics

Over the past three years, USDA has made substantial progress towards its goal of being a facts-based, data-driven organization. This includes establishing the CDO role with the responsibilities required under the Evidence Act and creating the Assistant Chief Data Officer (ACDO) role in each USDA Mission Area, supported by centralized analytics teams to perform and advance the use of analytics. The Department has also developed enterprise-wide dashboards to support improved decision-making while reducing manual data collection, as well as a common data and analytics platform that aggregates more than 3 billion records and 150 data sources from USDA agencies and other external sources.

In 2018, USDA successfully deployed enterprise administrative dashboards that provide leadership with insight into common functions such as Human Resources, Information Technology, Finance, and Fleet. In 2019, the Department partnered with all eight USDA Mission Areas to improve program delivery and customer experience through a focus on data and analytics. Today, thousands of employees across USDA have been trained on the use of data visualizations, resulting in a significant increase in the use of analytics products across all levels of the organization. Over 10,000 leaders, from the Secretary's office to front line leaders in nearly every county in the United States, now have direct access to data with a click of a button.

Achieving USDA's Vision

USDA's accomplishments over the last three years have dramatically improved access to key data and analytics products. Looking ahead, USDA will focus on how to scale and sustain its capabilities to fully embed data and analytics into how the organization operates every day, resulting in enhanced and more efficient service delivery for its customers.

> The following strategic goals of USDA's data strategy for FY21-23 will continue driving USDA's data-centric vision forward, enabling activities to better serve customers across and outside of the Department.

Through execution of this data strategy, USDA will streamline the ways in which it makes data and insights available to a broad set of stakeholders, increasing accessibility and scalability to effectively meet the varied needs of agency customers while providing direct access to meaningful, customized information from multiple data sets and research partners. Achieving this vision will allow USDA to solve cross-agency questions by accessing and joining data sources from across the Department on a common analytics platform, supported by maturity in governance, infrastructure, analytic capabilities, and data sharing. In the next year and beyond, the Department will continue to focus efforts on integration, collaboration, and sharing data to enable collaborative research capabilities and advanced analytics, promoting transparency across Mission Areas and providing customers with self-service functionality to easily access more data through fewer clicks, at a scale that meets their needs.

USDA FY21-23 Data Strategy Strategic Goals

- Goal 1-Data Governance and Leadership: USDA will strengthen data governance and data leadership to enable a strategic approach to data and analytics development, infrastructure, and tools.
- Goal 2-Data and Analytics Workforce: USDA will create a strong, data-driven culture by recruiting, retaining, and retraining the workforce to acquire the needed data and analytics skillsets.
- Goal 3-Common Data and Analytics Platform: USDA will develop and leverage technology, infrastructure, and analytics tools to enable shared access and use of data to achieve our mission and to drive innovation.
- Goal 4-Open Data: USDA will support and promote effective data sharing to provide customers, stakeholders, and the public with deeper insights, value, and transparency.

Goal 1: USDA will strengthen data governance and data leadership to enable a strategic approach to data and analytics development, infrastructure, and tools.

Data Governance and Data Leadership

In 2019, USDA completed a Data Management Maturity Assessment (DMMA) using the Capability Maturity Model Integration (CMMI) Instititute's Data Management Maturity Model to analyze USDA's eight Mission Areas against best practices and perform a gap analysis of the overall governance and management of data.

USDA has implemented several key recommendations that were delivered from this assessment including:

- Establishment of a USDA Data Governance Steering Committee, which sets and enforces priorities and policies to leverage USDA's data as a strategic asset, promote data stewardship, data sharing and efficient use of resources, and enable mission delivery.
- Establishment of the Assistant Chief Data Officer role in Mission Areas to lead data governance, strategy, and analytics activities aligned with departmental data strategies
- Finally, the Department established a department-wide data and analytics platform, the Enterprise Data Analytics Platform and Toolset (EDAPT), that provides a standardized, centrally available set of tools and connected data sources to enable analytics from descriptive to advanced predictive and natural language processing. EDAPT offers a single, common platform that eliminates technology as a barrier to sharing data and analytics products within and across agencies, and integrates data for analysis from more than 150 data sources brought in from every corner of USDA as well outside external sources.

Assistant Chief Data Officers

USDA established the role of the Assistant Chief Data Officer (ACDO) within each Mission Area to lead Mission Area data strategy and governance activities, and provide leadership in the following areas:

- Establish consistency and governance in managing data across the Mission Area so that data can be more easily used and shared within proper security parameters
- Oversee centralized analytics teams and provide analytics capacity to address cross-cutting questions or issues within the Mission Area
- Enable program areas to conduct program-specific analytics with common tools
- Ensure Mission Area alignment with USDA data strategy

> During FY21-23, USDA will mature the ACDO function, developing deep expertise in data analytics and governance. ACDOs will drive implementation of their Mission Area data strategies and oversee their Data Governance Boards in making strategic decisions, in addition to working closely with data stewards to accomplish new milestones in support of the USDA Data Strategy and Federal Data Strategy.

To ensure that the business needs of the Department are supported by the proper data, tools, and platforms, a critical component of ACDO responsibilities is to coordinate with Assistant Chief Information Officers (ACIOs) as counterparts in ensuring that the departmental vision is carried out at the Mission Area level and that business and technology requirements are met to support this vision.

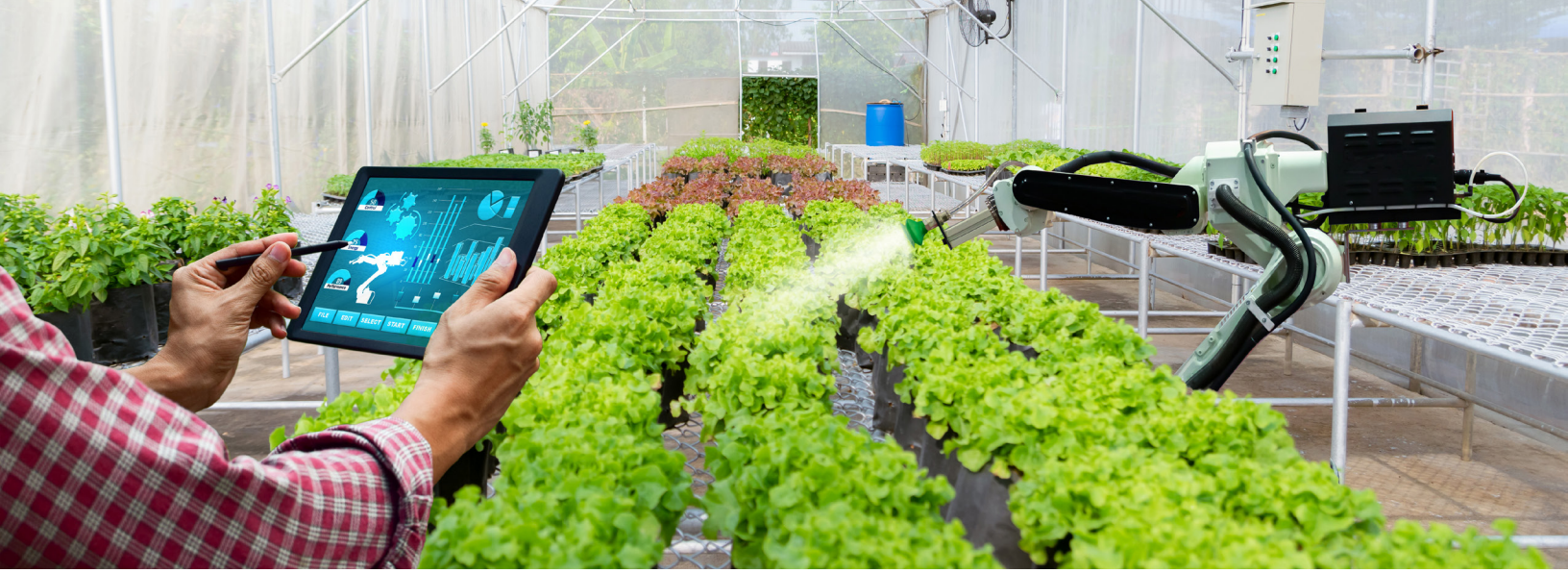
Data Stewardship

A data stewardship framework was established in FY20 to enable data stewards across USDA to develop data standards, data quality measures, and to collaborate on cross-cutting data issues with the goal of increasing the value, quality, and utility of USDA data.

> In FY21-23, data stewards will continue to play a critical role in the strategic initiatives supporting the USDA Data Strategy, including developing data quality standards, implementing a data catalog, supporting adoption of new tools to enable data sharing across the enterprise, and prioritizing data assets to make publicly available.

Dashboard Governance

USDA has implemented governance around the creation of dashboards to ensure efficient development of consistent data products on cross-cutting questions and wide sharing of analytics products across the Department. Dashboards reduce the manual collection of data, provide more timely access to information, and provide more intuitive insight into vast volumes of data.



Goal 2: USDA will create a strong, data-driven culture by recruiting, retaining, restructuring, and retraining the workforce to acquire the needed data and analytics skillsets

In order to fully leverage USDA's data as a strategic asset, the Department must ensure that its general workforce has the needed data skills and acumen to support a data-driven organization, expanding data expertise beyond data-oriented positions (e.g. data engineers and data scientists). Several components within USDA have well-established data programs with effective data quality standards, and the Department recognized opportunities to leverage best practices across the organization for improved data quality, documentation, and data sharing. The Department also recognizes that staff across USDA need additional competencies in data stewardship, basic data acumen, data visualization, and advanced analytics skills.

> For FY21-23, the Department will focus on building out the capabilities of existing staff, in addition to strategically acquiring and placing new talent to develop a skilled and knowledgeable, data-driven organization.

Centralized Analytics Teams

USDA has developed an organizational model for centralized analytics teams in each of its Mission Areas. These teams, led by the ACDO, support their respective Mission Areas by answering key cross-cutting questions with data while fostering analytics training and development more broadly across the Mission Area. USDA will ensure that each Mission Area's centralized analytics team is sufficiently staffed to support its size, needs, and maturity.

Centralized analytics teams provide support to their respective Mission Area ACDO by performing the following activities:

- Identify and help solve major cross-cutting strategic questions through the use of data analysis and advanced data analytics techniques and methods (including advances in data science such as machine learning, neural networks, and other forms of artificial intelligence)
- Create analytics products such as data visualizations, scenario analysis tools, and prescriptive or predictive models to draw insight from across Mission Area data sets — structured, semi-structured data, and unstructured (e.g. text) — for day to day use by business leaders
- Provide ad hoc analytics services to various parts of the Mission Area
- Cultivate a data driven organization through the development and enablement of the workforce

Workforce Skills Assessment and Development

> In FY21, USDA will conduct an assessment on data skills needed for the broader workforce to enable the Department to fully leverage its data as a strategic asset. The assessment will assist the Department in identifying knowledge gaps and educational needs and provide a roadmap for closing skills gaps by the end of FY23.

As part of the workforce development effort, Mission Area leadership will develop standardized, comprehensive job descriptions to create additional positions where gaps cannot be filled by existing staff, ensuring consistent and accurate hiring to close staffing gaps across all Mission Areas. Getting the right people in the right positions will be a crucial step toward enabling a more efficient and effective workforce to enact the priorities laid forth in published Mission Area, Department, and Federal data strategies.

Fostering Communities of Sharing and Education

Analytics tools and infrastructure are only truly valuable with informed and empowered staff, and USDA is committed to investing in the knowledge and problem solving acuity of its workforce to support each other in fostering a high-performing, strategic, data-driven community to promote the continued development of effective analytics products. To address the need for greater integration and to foster a more collaborative, data-driven environment, USDA established two communities of practice (CoPs) under leadership of ACDOs, bringing together individuals with an interest or established skill set in data visualization and analytics. Through these CoPs, best practices and tips — as well as examples of advanced dashboard development and data usage — are shared, encouraging questions and conversations among the groups and effectively raising the collective awareness and knowledge levels of cross-departmental staff.

- The Data Visualization CoP promotes the greater adoption of data visualization tools to assist all Mission Areas in building a culture that values data and promotes public use. The community also focuses on governing, managing and protecting data, as well as the most efficient and appropriate uses.
- The Advanced Analytics CoP is a community of analytics professionals that share research and advice and address questions. The CoP unearths and showcases institutional knowledge and experience in relation to analytical approaches to best practices, methodologies and tools and cultivates a data-driven organization through the development and enablement of the workforce.

Sharing best practices, challenges, and successes is proven to be an effective way to improve data management and analytics throughout the Department by fostering an openly collaborative learning environment.

Goal 3: USDA will develop and leverage technology, infrastructure, and analytics tools to enable shared access and use of data to achieve our mission and drive innovation

USDA strives to provide all users access to a common platform with standardized tools in order to enable the consistent development of analytics products at the Program, Mission Area, and Enterprise levels. The Department's broad array of data analytics platforms and tools consists of legacy technologies that are duplicative, difficult to maintain, and perpetuate data siloes. To achieve the Department's vision, USDA has established the Enterprise Data Analytics Platform and Toolset (EDAPT), which facilitates collaboration and data sharing and provides access to a standardized set of data science tools in support of the development of timely, accurate, and scalable analytics products.

> To continue the Department's recent progress, USDA's enterprise analytics platform modernization efforts over the next three years include several key areas of focus:

- Implement a single, common USDA data warehouse and analytics platform that eliminates data siloes and brings together data from disparate sources to enable secure data sharing and unlock analysis of important cross-cutting questions
- Establish a standard analytics toolset and reduce the number of analytics tools to enhance collaboration across the organization, while decreasing costs and driving administrative efficiencies through a centralized approach
- Develop a common data cataloging tool that provides visibility into the full scope of data across USDA and allows analysts to quickly and securely find and leverage data sets throughout the organization
- Develop a Data Science Workbench that provides a high-powered suite of cloud-based analytics tools, allowing analytics practitioners across USDA to fully unlock the value of their data through advanced analytics, data science applications, and collaboration on analytics problems
- Create an open data platform to enhance the way that USDA shares its data and analytics products with the public
- Develop an integrated strategy for how geospatial data can best be leveraged across the Department for improved decision-making

Data Warehouse

To support its enterprise analytics data sharing capabilities, USDA is establishing a single, common, enterprise data warehouse platform that modernizes and consolidates individual Mission Area and component agency data warehouses and provides secure, segregated data storage and access for each Mission Area. The EDAPT platform already provides the underlying structure to enable users to share data securely across the Department, ultimately allowing programs to better address important program questions collaboratively.

> Slated for completion in FY23, the enterprise warehouse will serve as a crucial part of the departmental infrastructure by promoting security, integrity, and centralization for data storage and management to support analytics capabilities.

Enterprise Analytics Tools

With a centralized analytics platform, USDA has the opportunity to standardize on common analytics tools, which will vastly increase the ability of analytics practitioners across the Department to collaborate on common or cross-cutting challenges and to share results with end-users in a single, user-friendly environment. Additionally, this standardization will enable the Department to centrally administer and license these tools, leading to additional cost efficiencies in how USDA extracts value from its data.

USDA standardized on an enterprise data visualization tool in FY20, providing end-users with access to dashboards at the Department, Mission Area, and program area levels in a single pane of glass with the click of a button.

> In FY21, USDA will launch a Data Science Workbench that provides an expandable suite of tools for performing analytics and data science processing (e.g. Python, ESRI, etc.), enabling agencies to perform complex analytics in a scalable cloud environment.

The Workbench allows users from across the Department to harness data from disparate sources to address questions on program performance, evaluation, and research, improving business outcomes and decision making. The Data Science Workbench offers easy, widespread access to standardized tools in an expandable suite, facilitates learning and adoption of analytics practices, and empowers leaders and employees across USDA to become more data-driven.

Data Catalog

To provide governance over the data and to ensure standardization, transparency, and consistency of various sets of program data being shared and consumed, USDA is establishing an enterprise data catalog and data dictionary.

> In FY21, data stewards and other knowledgeable data stakeholders will compose an inventory of their program area's data with key attributes. The implementation of USDA's data catalog will mitigate common issues shared across the organization related to difficulty finding and sharing data sets across Mission Areas.

Geospatial Data

The USDA CDO works closely with the Geospatial Information Officer (GIO) and members of the geospatial community, soliciting input and guidance on GIS initiatives and ensuring coordination on activities related to geospatial tools and capabilities. As geospatial technology matures within USDA, Mission Areas such as Farm Production and Conservation (FPAC) are already leveraging it for uses such as Artificial Intelligence (AI)/ Machine Learning (ML) land change analysis, disaster assessments, and locational analytics to understand optimal field office locations.

> In FY21 USDA will publish a Geospatial Data Strategy in alignment with the requirements of the Geospatial Data Act (GDA).

Interdepartmental initiatives, such as the USDA/NASA umbrella and annex agreements, leverage geospatial agreements to increase the amount of shared data collected by multiple agencies, reducing customer data requests and leading to cost savings and efficiencies for the organizations. In line with this goal, the GDA aims to expand Department-wide access to geospatial data sets that can be managed as a shared service, and capitalize on opportunities for shared infrastructure and information technology (IT) services.

Streamlining Data Collection Methods and Common Customer Records

USDA's agencies each pursue distinct missions to support the agricultural community, made possible by the identification, collection, and applied use of research data specific to their needs; however, there are many overlaps in needs for specific data sets. USDA has opportunities to continue reducing the burden on customers and provide timely, accurate data, which is often manually collected from farmers and ranchers through surveys. For this reason, USDA will explore alternative methods by which data can be collected and shared for analysis, to the extent permitted by law and with appropriate privacy measures in place. Instead of requiring that farmers and ranchers provide the same data, multiple times, in various formats, to different agencies, USDA is working to ensure that data can be gathered only once and shared accordingly to provide information back to customers in meaningful ways.

USDA will work to reduce the burden on farmers and ranchers in the data collection process by leveraging new sources of data, such as geospatial data and computer vision. USDA will continue to provide measurable value back to customers through analytics products developed from the data collected. As an example, the National Agriculture Statistics Service (NASS) is developing a tool that will leverage multiple data sources and integrate existing systems (e.g. working with the Farm Service Agency (FSA) to connect Farmers.gov data) to reduce process steps and timelines for data collection and reporting. With this tool, multiple customer records from various sources can be shared, stored, and provided to customers, reducing the amount of data requested of farmers and providing customized views of data.

Goal 4: USDA will support and promote effective data sharing to provide customers, stakeholders, and the public with deeper insights, value, and transparency.

USDA is home to an immense resource of agricultural data, and is committed to creating an open data strategy that supports researchers and the public in accessing and using data. With public-facing data, USDA can reach its customers directly, enabling interactive, direct access to these vast data sets and delivering valuable insights through visualizations. This aligns with USDA's objectives to improve transparency and provide the public with access to information that is not otherwise restricted by governing laws and regulations. Increased integration and analysis of this agricultural data is also key to supporting innovative analyses by university researchers and is essential to addressing some of the most pressing questions facing the U.S. agricultural sector and American society at large.

Open Data Strategy

USDA will develop an open data strategy in FY21 to accomplish the following priorities:

- Identify USDA's priority data assets for its open data plan, describing data assets that are especially valuable to the public interest and therefore will be made available as open government data
- Develop the Department's first open data plan, as required by the Evidence Act
- Provide data to inform local and national agriculture policy development
- Protect the confidentiality of producers and agricultural business data
- Abide by relevant laws and regulations
- Facilitate improved public understanding of USDA programs
- Support research efforts on agricultural production and conservation

USDA contributes over 1,800 data sets to the federal data catalog on data.gov but more can be done to make this data accessible and insightful. The establishment of a USDA open data platform reflects the Department's commitment to facilitate public access to useful data in an easily consumable format.

> In FY21, USDA will launch its Open Data Platform, enabling efficient publication of dashboards and data to public websites. The Department will also continue to enable more direct data access, empowering stakeholders to leverage open data directly through application programming interface (API) connections.

USDA regularly partners with other federal agencies, state partners and other stakeholders to share data and information to support federal programs. Examples of this include the Animal and Plant Health Inspection Service (APHIS) who partners with State veterinarians on animal health matters, the Forest Service who partners with the Department of Interior, States and other organizations to fight wildfires and the Food Safety and Inspection Service who partners with the Food and Drug Administration and Centers for Disease Control on foodborne illness outbreaks. These partnerships are limited in their ability to share data due to separated networks and siloed collaboration solutions. The USDA Open Data Platform also supports these partnerships through a common data platform with controls to ensure appropriate access when needed.

USDA also works with researchers and academia to provide technical assistance, program evaluation and scientific insights to further USDA's mission. USDA intends to enhance its ability to share data in a secure environment so that important analytics and research can be performed with the support of expertise from research institutions.

Engagement with other Federal agencies, research partners, farmers, ranchers, and customers is critical to improving access to USDA's data that can answer critical research questions. Leveraging a strong underlying structure of data governance and data stewardship principles, as well as modernized technology platforms, USDA is positioning itself to support these mutually beneficial partnership opportunities while ensuring the privacy and security of its customers.

Conclusion

Through the work done to date, USDA has ensured that the Department's stakeholders and data users – both internal and external – have the infrastructure and technology available to receive the data they need when they need it, share data to promote research and learning, and integrate data to develop advanced analytics products.

Through its data strategy, USDA will continue to strengthen data governance and data leadership to enable a strategic approach to data and analytics development, infrastructure, and tools. The Department will also create a strong, data-driven culture by recruiting, retaining, and retraining the workforce to acquire the needed data and analytics skillsets. To remove technology as a barrier, USDA will continue to develop an enterprise analytics infrastructure and toolset to enable cataloging, integration, and sharing of USDA's vast data across Mission Areas, improve internal decision-making and efficient use of resources, and provide the public with easy access to data that can drive commercial innovation.

Finally, USDA will prioritize data assets for its open data plan and effectively engage with the public and the private sector to support data sharing while providing value back to its customers with deeper insight into the data collected from them. By taking these key steps, USDA will become the best managed agency in the Federal government with a relentless focus on being facts-based and data-driven.

Office of the Chief Information Officer
U. S. DEPARTMENT OF AGRICULTURE