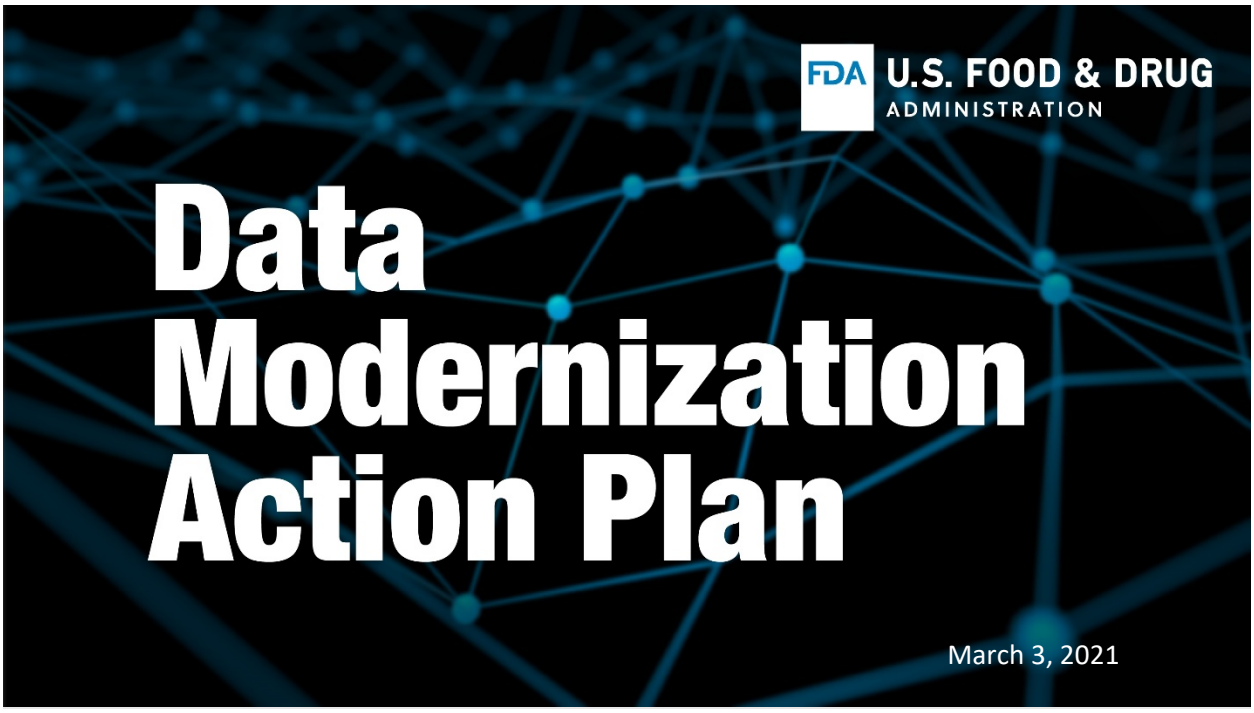


Introduction: Science-based regulation of evolving technologies will require FDA to develop new ways of interacting with data

Data have always formed the basis of science-based regulatory decision-making. These data may come from traditional sources—for example, measurements submitted to FDA from clinical trials or observations from FDA field inspections. As technology becomes more sophisticated and our world becomes more connected, data from many new sources can help us understand how medical products are performing, pinpoint the source of foodborne illness, or understand emerging public health threats.

At the same time, FDA's regulatory mission is growing more complex with the technological evolution of FDA-regulated products. New technologies hold enormous promise to patients and consumers—and unique challenges for responsible regulation. Without modern expertise and approaches to managing and analyzing data, FDA will miss critical opportunities to benefit patients and improve public health.

Data Modernization at FDA

In today's world, most interactions and processes are instrumented and digitized creating an abundance of data. Leveraging this data using modern data techniques will unlock new insights and value for public health.

Decades ago, much of the information submitted to FDA was not in digitized form—think of a hand-drawn graph representing the observations of an experiment. However, as digital technologies have become the norm over the past few decades, our society's ability to capture, analyze, and display data has created opportunities to analyze these data in powerful new ways.

And yet, FDA's data systems are still largely geared to a non-digital, document-based information paradigm. Electronic documents (e.g., PDFs) are not the same as digital data. Although FDA's legacy technology and data systems allow the Agency to meet its regulatory responsibilities, FDA urgently needs new, robust, and flexible capabilities to avoid losing future opportunities.

As part of FDA's data modernization, we are developing a responsible strategy for deploying new data systems to allow FDA to manage and analyze data that previous generations of researchers, physicians, and FDA reviewers could have only imagined. During the course of this work we will ask: Where can new methods, such as artificial intelligence, be applied to increase the effectiveness of FDA reviewers and field investigators while ensuring analytical rigor and avoiding bias? How can we develop flexible data tools that can be deployed across FDA's diverse program areas to reduce resource demands? How can we manage and share data in a way that ensures privacy, security, and transparency?

FDA's data modernization is the next step in the Agency's overhaul of its approach to technology and data

The Food and Drug Administration (FDA or Agency) has multi-faceted responsibilities – protect the public health by regulating medical & veterinary products, food, cosmetics, and tobacco products, speed innovations that make products more effective, safer and more affordable, and ensure the security of the food and medical supply chain by responding to emerging public health threats.

These responsibilities require the FDA to continuously monitor trends in science, technology, and data for their impact on public health and to improve the Agency's own efficiency and effectiveness.

To address these objectives, the Technology Modernization Action Plan (TMAP)[1], published in September 2019, laid out a foundation for modernizing the FDA's technical infrastructure, and signaled the development of the FDA's ongoing strategy concerning data management and data itself. Specifically, this framework for FDA's Data Strategy will focus on the stewardship, security, quality control, analysis, and real-time use of data to accelerate the path to better therapeutic and diagnostic options for people and animals, better secure the food supply chain, and provide state-of-the-art tools to enhance and promote public health.

FDA's Data Modernization Action Plan

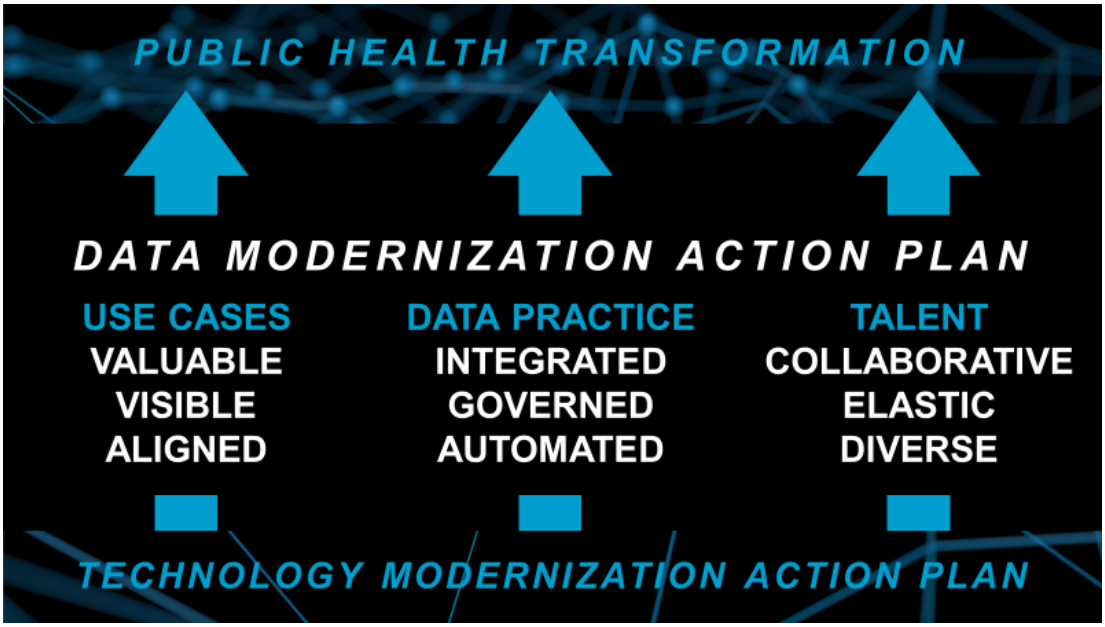
The Data Modernization Action Plan (DMAP) proposes a framework and actionable recommendations for FDA's Data Strategy. It consists of 3 key components:

(1) **Identify and execute high value driver projects for individual centers and the Agency;**

(2) **Develop consistent and repeatable data practices across the Agency; and,**

(3) **Create and sustain a strong talent network combining internal strengths with key external partnerships**.

[1] https://www.fda.gov/about-fda/reports/fdas-technology-modernization-action-plan



I. Why do we need a comprehensive data action plan now?

Complexity and Interdependence: Evolving technology is revolutionizing human and animal health. These new technologies feature increased complexity of scientific and process-oriented data, as well as the opportunity for personalization and increased precision when addressing:

- Data from multiple touchpoints of public health, as they are not independent silos but essential components of the totality of information.
- The processes supporting public health are highly interdependent. Issues or changes in one process have ripple effects across the public health ecosystem.
- Technological advances that continue to test the boundaries between product types and the historical lines of separation between FDA's various product centers; shared capabilities can be achieved through shared data.

FDA will design and execute a framework for FDA's Data Strategy that reflects this environment of interdependence. The strategy will also align with Federal government programs, policies, and statutory requirements.

New Application of Available Data and Technologies: The FDA's role in regulating food, tobacco, cosmetics, and human and veterinary medical products gives it a unique responsibility to align with technological change and innovation across these industries. Digitization of processes, pervasive use of mobile technologies, and easier access to computing resources have created new types of data. These data types and related technologies can be used to create innovative solutions to improve public health, such as:

- Capabilities to track and trace medical and food products can transform national response to emergencies by identifying product and logistic information across the entire supply chain.
- Integration of real-world and clinical trial evidence can increase representation of diverse populations. It can transform overall efficiency of product reviews and post market surveillance.

- Privacy sparing innovations can advance treatment of rare diseases while protecting appropriate patient specific information.

> **Example Innovation and Partnership**: FDA is an active participant in the COVID-19 Evidence Accelerator,[2] which is helping to answer critical questions related to COVID-19 diagnostics and therapeutics using real world data and modern, rigorous analytical methods.

FDA will partner, where appropriate, with patient groups, consumer groups, technology companies, non-profits, industry partners, academia, and other stakeholders to create platforms to accelerate innovation.

Scaling the Agency: The FDA's volume of work is increasing markedly, requiring utilization of new technology and complex datasets to increase efficiency. As a result, talent elasticity should be improved to meet unexpected demands and minimize impact on planned activities.

- FDA's portfolio of work is rapidly expanding due, in part, to the addition of new classes of medical products to the FDA portfolio.
- Applications for medical products, such as combination products, novel medical devices, and orphan drugs, will increase the demand for even more resources.
- FDA must prioritize and screen an increasing number of import lines to ensure our investigators can perform the work needed efficiently and effectively.
- Responsiveness to unplanned events such as pandemics, foodborne illness, natural disasters and global events affecting the supply chain requires coordination across multiple specialties and global entities, stressing the same resources used for day-to-day regulatory activities.

> **Example Scaling**: Advanced Semantic Search and Indexing of Text for Tobacco (**ASSIST4Tobacco**) pilot to explore and build a model to scale search capabilities for complex Tobacco submissions. Improved search capabilities based on context and intent will scale review, monitoring, and regulatory operations by reducing the burden of manual, inefficient process on FDA staff.

FDA must increasingly use solutions such as artificial intelligence (AI) and other automation techniques to adapt to changing priorities and leverage our staff in a collaborative and elastic manner. A robust data strategy is critical to meet this goal.

II. How will FDA pursue Data Modernization?

Element 1: Strong Driver Projects: The DMAP is anchored on driver projects that help generate value while building critical capabilities. Driver projects for DMAP are defined *as initiatives with measurable value that help multiple stakeholders envision what is possible, allow technical and data experts to*

[2] https://www.evidenceaccelerator.org/

identify needed solutions, and develop foundational capabilities. This strategy is distinctly different from focusing on data collection and then looking for questions the data can answer.

The selected driver projects will not only address traditional performance indicators but will also support transformation across the Agency by using predictive models, and appropriate application of trends such as AI. The use cases can be anywhere in the continuum of a center-specific opportunity to an Agency-wide capability. Some examples of driver projects:

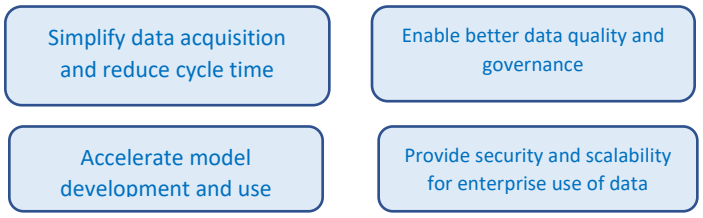
- Identify and incorporate lessons learned from the AI seafood pilot for import inspections to scale the capabilities for other food and medical imports
- Use internal and external data to pinpoint critical points in the supply chain that could lead to drug and food shortages due to regional, national or global emergencies

ACTIONS

- Develop an Agency-wide framework to identify, prioritize and execute driver projects
- Leverage each driver project to deliver value and contribute to the maturity of data practices
- Measure and communicate value of driver projects to internal and external stakeholders to support the Agency mission and promote transparency regarding data capabilities

Element 2: Data Practices:

The driver projects-based approach allows the Agency to show near-term value in specific areas while enhancing key data practices. However, a modern data strategy also requires proactive investments in foundational capabilities. For example, reducing time to perform repeatable tasks such as identifying the right data for a given problem (currently 50%-75% of the total time spent) will greatly reduce cycle time and better leverage precious analysis resources. Data practices should achieve the following goals:



Key components of Data Practices to achieve these goals are **Identification**, **Data Curation**, **Governance and Automation**

Identification: Building effective solutions using data requires the ability to identify the right data with the appropriate quality and completeness. Example capabilities are:

- Data catalog to search and review available data assets and relevant metadata.
- Workflow process to request and approve access.

Data Curation: Data curation is the organization and integration of data collected from various sources for effective utilization in data driven decisions. Curation increases confidence in use of data and reduces cycle time without compromising the analytic rigor. Key activities and capabilities for data curation are:

- Characterization of data on attributes such as latency, completeness, security/privacy and quality.
- Classification of users based on proficiency, and analyses based on impact of decisions.
- Development of "Fit for Purpose" services to match the curation with analytic needs using an agile and scalable model.

Governance: Governance in the DMAP context spans the entire lifecycle of data and appropriate upstream and downstream activities. The objective of Governance is to reduce waste and friction in using data for decisions. The Governance program should be expanded to include appropriate use of models and AI based algorithms, including:

- Technology, tools, data policies, analytic techniques, information disclosure and security processes
- Reproducibility of results used for research
- Architecture of transaction systems to reduce the need for curation

Automation: To meet the exponential growth of data and its pervasive use across the Agency, many data capabilities should be automated to assure reliability, predictability, and timeliness of data driven decisions. The Agency will:

- Learn from automation of application and security processes at the Agency in developing Data Operations.
- Place special focus on changes, prevent the introduction of potential bias and assumptions in data that drive models and algorithms.

> **Example Capability**: The Global Substance Registration System (GSRS), designed by the FDA, provides consistent, auditable, quality and uniquely identified substance information for reviewers and scientists. Attention to data management practices such as use of international data standards, expert curation, stewardship and program area governance ensures that this foundation data is useful for all stakeholders.

ACTIONS

- Conduct an assessment for key data practices
- Establish an Agency-wide end to end Governance model
- Partner closely with application teams to build architectures that can enable easier data analysis
- Adapt automation techniques from software development and other practices
- Pilot an initiative to create an Enterprise Data Model for the Agency

Element 3: Talent Development:

Without a strong focus on talent and elastic talent networks, the modernization plan will be slow, inconsistent and expensive. Several critical factors need to be considered as part of a talent strategy, including:

- Compensation: strong competition from private industry with different pay scales.
- Specialized data science skills are in demand and far outweigh the available talent pool.
- Data skills are fundamental, and experts are cross-trainable to increase overall capacity.

Skills for successful data science practice: in addition to data science experts, there is need for domain experts, data modelers, computer and software engineers, statisticians, software engineers, data modelers, story tellers, product managers, and agile practitioners.

Example of talent flexibility: FDA developed an Agency-wide advisory matrix for surveillance inspections in less than six weeks by using data available from HHS and combining it with internal data. A small team of specialists and volunteers used agile methods to complete the project and launched it for all centers and State government liaison officers

The FDA should create specific solutions catering to its staff and unique needs. It should also leverage existing programs across the Federal government to addresses these challenges.

ACTIONS

- Identify critical skills for executing successful data projects
- Conduct internal and market surveys of effective methods of training
- Develop strategies for recruiting and retaining talent
- Create an internal network to share knowledge and resources
- Pilot and scale a Data Science platform for Agency-wide use

Develop a nimble operating model to execute driver projects and scale it for broader use.

III. How will FDA measure and communicate the progress of the DMAP?

FDA prioritizes transparency and accountability in the implementation of DMAP. Ongoing evaluation of our progress will enable iterative improvement of our modernization strategy.

- **Establish value based measures**: Measures to demonstrate cycle time reduction, broader sharing of data and capabilities across the Agency.
- **Engage stakeholders**: FDA will engage stakeholders at multiple points during this process, to solicit proactive ideas and reactive feedback.
- **Provide annual updates**: FDA will provide public, annual updates on DMAP-related activities, including accomplishments, challenges, and lessons learned.

IV. What comes next?

DMAP will leverage the foundations laid by TMAP (modern technical infrastructure, a product-oriented approach, collaboration) such as Cloud platforms and product mindset to execute DMAP. The two action plans will go hand in hand to realize the full potential and value of data and technology for the FDA and its stakeholders.

Near term activities for DMAP execution:

- Form an Agency-wide Steering Committee to refine and implement DMAP

- Build on the experience of TMAP execution and Center experiences to begin planning for resource needs
- Conduct discovery sessions to identify best practices within and outside the Agency
- Finalize the execution plan and identify driver projects
- Seek feedback on the plan and make necessary refinements to execution
- Launch specific programs mentioned in the Actions section for all three components