USAFacts CODE

Building the National Data Ecosystem America Deserves

A Report from USAFacts and the Center for Open Data Enterprise (CODE)

Table of contents

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

Executive summary

Executive summary

> **A popular Government, without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or, perhaps both. Knowledge will forever govern ignorance: And a people who mean to be their own Governors, must arm themselves with the power which knowledge gives.**
>
> James Madison, writing in 1882[1]

From the beginning, America's founders recognized that data, information, and knowledge are critical to a healthy and robust democracy. The country's commitment to national data and statistics has grown from the first U.S. Census, launched by Thomas Jefferson in 1790, to a sprawling, complex collection of hundreds of thousands of federal data sources.[2] This diverse, decentralized ecosystem produces data that is essential for Americans to understand the country, assess our progress, and govern effectively.

Now the U.S. federal data system is at a point of transition. Evolving politics and changes to the federal government have led to discontinuing some data collections, major staffing changes in data-producing agencies, and proposed alterations to the Census and other key data sources. A number of significant federal data sources have been lost, and several organizations are tracking the system to identify other datasets that may be at risk (see Appendix 6).

At the same time, the need for data innovation has grown as traditional survey instruments have become less reliable. The rise of artificial intelligence (AI) and mass digitization of records are providing new opportunities for data generation, analysis, and use, but also raising concerns about data quality, accuracy, and privacy. It's critical for America's data systems to evolve and improve to meet the changing needs of the population and the times.

These disruptions have created a unique opportunity. We are at a moment when we can reimagine America's national data ecosystem holistically, in a way that centers on core principles while ensuring the flexibility necessary to remain relevant in the face of change. It's an opportunity to rethink a system that has grown without an overarching plan for more than two centuries, and move to a system shaped more intentionally to meet national needs.

What's more, there is urgency to reimagining America's national data ecosystem. As technology advances rapidly with the growth of AI, the national data ecosystem risks falling further behind. The result could be a greater disconnect between government data — foundational information on which Americans rely to understand the state of their country — and the technology that they use to answer their most pressing questions, threatening trust in government and further dismantling a shared source of knowledge that underlies public discourse.

The organizations USAFacts and the Center for Open Data Enterprise (CODE) are publishing this paper to help lay out a vision for that future data ecosystem.[3,4] The paper is based on two Roundtables on Building the National Data Ecosystem America Deserves that were held in Washington, DC and Oakland, CA.

The Roundtables followed CODE's established convening methodology[5] and engaged a total of 154 stakeholders. They included members of civil society, research and academia, community-based organizations, business and business associations, state and local government, and former federal government employees. (Appendices 2, 3, and 4 provide a list of participating organizations and the Roundtable agendas.) USAFacts and CODE worked with an advisory group of 10 experts, conducted interviews with 15 others, and analyzed hundreds of pages of notes and transcripts to identify major themes from the Roundtables (with some help from Generative AI).

Through the Roundtables and related efforts, we found strong agreement about the kind of national data ecosystem that would best serve the national interest. This paper presents our findings as a starting point for the work ahead. It is not meant to be a final, complete analysis of all the issues impacting national public data, but an overview of high-profile issues and opportunities that emerged from the Roundtables and our research. We recognize that national data serves a wide variety of stakeholders who may have different perspectives and data needs. These include:

- Members of the public who benefit from data (even if they are not aware of it)
- Nonprofits and NGOs working toward mission goals
- Community-based organizations
- Scientists and researchers in and outside of academia
- Members of the media
- Businesses of all types and sizes, including large tech companies and smaller startups that use data and AI applications
- Federal, state, and local government data stewards, legislators, and policymakers, who may be both data providers and data users

We also recognize that dozens of organizations, ranging from longstanding institutions to those launched in the past year, are now doing essential work to strengthen national data at this critical time. Our online resource, the National Data Ecosystem Resource Hub, provides a guide to the many organizations who participated in the Roundtable.[6] We welcome additions to this resource and hope it will be a useful tool to facilitate collaboration on common goals.

This paper presents our findings in three major sections:

1

Envisioning the national data ecosystem America deserves

This section summarizes the major qualities that diverse stakeholders would like to see an ideal national data ecosystem embody. Some other organizations, such as the Association of Public Data Users, have proposed similar frameworks.[7] We hope that the framework presented here, drawn from an unusually diverse range of expert input, will help shape goals for future work. The framework includes three major categories: core qualities desired in a national data ecosystem, ease of access and use, and ethical and responsible management.

2

The federal government's essential data roles

Our national data ecosystem includes a wide range of sources, including state and local data, private-sector data, community-generated data, and more. Federal data, however, has to be the strong foundation for the country's data and statistics. This section summarizes the roles that only the federal government can realistically play in ensuring robust national data.

3

Strategic and operational goals for an improved data ecosystem

This section draws on the Roundtable and other research to crystallize **seven strategic and operational goals** for a better, more robust national data ecosystem. It goes beyond general principles to describe concrete, measurable goals and ways they could be achieved.

These seven goals are:

Goal 1

Maintaining critical federal data collections

With federal data in a state of transition, it's important to identify essential data sources in different sectors, determine what data must be maintained, and develop criteria for deciding when and how a federal data program can be discontinued. There is also a need to monitor changes to government data collections and archive important data and tools.

Goal 2

Strong, reliable federal data governance

The current pressures on national data provide an opportunity to take a fresh look at how the government manages its data. It's an appropriate time to fully implement current laws and privacy protections, improve federal data products, and consider new legislation, governance structures, funding mechanisms, and technical solutions. Developing better ways to manage the country's data, and communicating them to the public, can help restore trust in our national data systems.

Goal 3

A national data ecosystem that serves all Americans

National data collections should represent all American communities and groups. It's important to ensure that data is disaggregated by race, ethnicity, income, education, rural or tribal residency, sexual orientation, and other factors at both a national and state level. It's also important to engage with local communities around that data collection and use.

Goal 4

An AI-enabled system for better data and data access

AI is having a profound impact on the federal data ecosystem. There is an immediate need to address critical concerns about AI while also improving national data in ways that will make it more useful for AI applications. At the same time, AI can improve the national data ecosystem by harmonizing data sources, creating new data sources, and democratizing data access and analysis.

Goal 5

Extensive and usable systems of state and local data

State and local data, an integral part of the national data ecosystem, needs to be improved in many ways. There are opportunities to improve capacity and ensure funding for state and local data, improve data sharing and data standards, preserve data for state and local use, and tap administrative data for analysis.

Goal 6

Robust American data developed with and for America's businesses

Both large and small businesses have a stake in a robust national data ecosystem, and they can do more to help build that ecosystem for the future. There are opportunities to explore using private data for public use and to identify public data of high value to the private sector. New approaches can include developing proxy and blended data sources and new cooperative models for data collection and management.

Goal 7

Shared expertise, communication, and collaboration around national data

National data is a critical public good, and the public can become more involved in its development, maintenance, and use. Stakeholders with an interest in national data can tell more effective stories, target key stakeholders more strategically, and build and tap new expert networks.

Achieving these goals will take vision, organization, collaboration, and time. That work will need to balance immediate improvements with more extensive initiatives to achieve bold goals for major change.

For each goal, we have listed both Ongoing Priorities and Emerging Opportunities.

The Ongoing Priorities may seem familiar: In many cases, they are needs for improvement that have been identified and debated for years. The fact that they are familiar issues, however, does not mean that they have been solved or that they should be ignored. With the current disruptions to America's data systems, this may be the right environment of change to surface these priorities and act on them.

In contrast, Emerging Opportunities are issues that our work has identified as new areas for significant improvement and change to the national data ecosystem. We hope it will be possible to act quickly on these in the current environment.

We hope that presenting these priorities and opportunities in this way will help organizations working in this field choose areas for action and collaboration to improve our national data.

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

INTRODUCTION

Government data at a turning point

Introduction: Government data at a turning point

> **This is a disruptive period. We want to take advantage of that disruption. We don't want to save what cannot be saved. We want to think differently and aggressively about how we can make new data and new statistics, a system that would be responsive and trusted, protecting privacy.**
>
> Bill Beach, former Commissioner, Bureau of Labor Statistics[8]

Americans interact with federal data every day without even knowing it. This massive trove of digital information impacts industry, public services, and daily life. Much of it is open data that can be freely used, reused, and distributed by anyone for any purpose.[9] Federal data is the basis for economic measurement and metrics, public health and educational programs, scientific studies, and more, and has fueled a wide range of local and national initiatives.[10] Americans interact with open federal data every day when they search the web, use their phones' GPS systems, or check a weather forecast. Since the country's early days, and especially over the past two decades, the U.S. government has set policies, passed legislation, and taken practical steps to improve the nation's data with virtually unanimous bipartisan support.

In principle, federal data is the best source for understanding what is going on in the country. It is funded by taxpayers, overseen by elected representatives, produced by some of the smartest and most dedicated statistical officials, and highly invested in — billions of dollars are poured into it each year. It is official and authoritative in an information ecosystem where very few sources are highly trusted. And the federal government is the only institution that can reliably produce certain datasets, including in situations that require significant funding, that are related to federal programs, or that are compiled from our nation's more than ninety thousand governments.

Because federal data has unique value, data users have become concerned about current disruptions to the federal data ecosystem. (A March 2026 report from the Congressional Research Service, *Availability of Federal Data*, provides a comprehensive view of current issues in the context of the systems governing public data.) In a survey before each Roundtable, USAFacts and CODE asked participants, "What challenges and opportunities are you seeing as a result from recent changes to federal data programs?" The dozens who responded flagged a number of issues, including constraints on resources in the form of funding, staffing, and technical capacity, the risk of political issues affecting data, and the erosion of trust in national data.

Many organizations that participated in the Roundtables are actively involved in archiving and preserving federal data, assessing current risks to the federal data system, advocating to support future data collections, or addressing disruptions to federal data in other ways. CODE has published an analysis of these issues in its report *America's Data Future* and focused on specific risks to public health data in a report on *Ensuring the Future of Essential Health Data for All Americans*.

But while these concerns were the context for holding the Roundtables, the Roundtables themselves were designed to address the future. The core question for these convenings was, "How can we leverage this moment of change to create something better than the system we've had in the past?"

The national data ecosystem has long been far from perfect. Many government data sources, platforms, and tools are not interoperable, are difficult to access or share, and are not ready to support AI applications. Our data infrastructure is dependent on legacy systems that go back decades and are often inefficient, hard to use, and outmoded. Low survey response and other issues have impacted the quality of a variety of data products, including flagship government surveys like the Current Population Survey. Different datasets may be duplicated or conflicting; some data sources are not timely enough; and quality and format issues make it difficult for AI applications to fully integrate federal data.

The challenges facing federal data stem partly from the nature of U.S. data infrastructure. The U.S. federal data system is not really a coherent "system" as much as a massive collection of programs run by many different agencies and offices. Where many other countries have national statistical offices (NSOs) that collect and publish statistics in a unified way, the U.S. relies on "an infrastructure of 13 principal statistical agencies (located within 11 larger federal 'parent' agencies) and over 100 other data-producing agencies throughout the U.S. government."

While U.S. data is governed by many individuals and organizations, including the U.S. Chief Data Scientist, the Chief Statistician of the U.S., and the Chief Data Officers (CDO) Council, there is no centralized accountability for the federal data ecosystem as a whole. Improving America's data will take leadership and a large-scale, collaborative effort across agencies and organizations.

Building a better data ecosystem will require not only protecting the integrity of high-value datasets today but also solving longstanding issues as we head into the future. We can identify the core data collections that have high value to the public and therefore should be sustained; improve those that can be managed more efficiently and effectively; sunset those that no longer serve a meaningful purpose; and prioritize data resources for preservation, development, and use. At the same time, we can look for additions to federal data collections, such as data from state and local governments and the private sector, where they are viable options.

Improving government data has never been more important. The emergence of AI is poised to reshape much of American society, including the labor market, the flow of information, and the pace at which technology is now expected to advance. To be a force for public good, AI will need both strong ethical and policy frameworks and timely, reliable data in formats that it can analyze. While the United States government data ecosystem is still considered the gold standard for the world, it is technologically and methodologically outdated in many ways, and is not up to the demands that AI will place on it. Modernizing the U.S. government data infrastructure is no longer just a wish: It is now a necessity to ensure that AI systems, and the public, have access to credible, reliable data about our world that can continue to inform our decisions.

USAFacts and CODE, each launched a decade ago, are among the growing number of organizations working to improve America's data ecosystem. CODE has a mission to harness the power of open and shared data for the public good by working with data providers, data users, and a wide range of stakeholders. USAFacts has a vision for a democracy rooted in trusted facts. This requires data that comes from the government; that helps policymakers make effective, transparent decisions on behalf of their constituents; that allows for evaluation of policy; that helps the public understand why they made these decisions; and that allows anyone who wants it the opportunity to understand the current state of the nation.

Achieving these goals requires a vision for a national data ecosystem that will best meet the needs of the country. The following section presents the qualities a future data ecosystem will need, as envisioned and defined by the organizations that engaged with this project.

BUILDING THE NATIONAL DATA
ECOSYSTEM AMERICA DESERVES

Envisioning the national data ecosystem America deserves

Envisioning the national data ecosystem America deserves

Roundtable participants identified a number of qualities that they believed are important to a strong national data ecosystem. They fall into three broad categories: the ecosystem's **technical foundation and infrastructure**, **the user experience** it should provide, and how it enables **public stewardship and social benefit**.

Technical foundation and infrastructure

An ideal national data ecosystem should be:

Standardized

Data standardization serves many functions. It can ensure consistency in data collections across political administrations; help make data interoperable across sectors (e.g., healthcare and housing data); facilitate data sharing; and help align government, academia, and industry around a cohesive data ecosystem. Data and metadata standardization is also essential for AI development.

> **Challenges**
>
> It's not clear who should design national data standards, who should enforce them, and what the consequences should be of disregarding them. The U.S. has a long and successful history of public-private collaboration for standards development.[11] Whoever develops these standards, whether the federal government, a trusted non-federal body, or a collaborative public-private group, will need to balance the value of stability in data collections with the need for adaptability to meet changing circumstances.

Machine-readable and AI-ready

The national drive to accelerate AI development demands better data to fuel AI analysis and applications. A national data ecosystem demands high-quality metadata, standardized formats, and widely used, widely accepted benchmarks that are regularly reviewed by humans. Additionally, the ecosystem needs to adapt as quickly as AI does, including testing and rolling out agent-to-agent communication protocols like the Model Context Protocol (MCP).

> **Challenges**
>
> Today's federal data ecosystem, with much data still held in PDF formats, is not even consistently machine-readable, let alone AI-ready. National data needs to be made available in standard formats with consistent schemas and metadata frameworks.

Resilient

Disruptions to federal data systems have highlighted the need to have a national data ecosystem that is resilient to rapid, unexpected data deletions or data program cancellations. Elements of resilience include:

- Methods to ensure data preservation and availability.

- Continuity and predictability across administrations. Core datasets and public access to them should be insulated from political pressure, leadership turnover, or shifting federal priorities.
- Decentralization, so that not all data is dependent on federal control. As one Roundtable participant put it, "When we build a hospital, we don't expect that the only source of power is the one that's coming from the grid. We have generators and backups."[12]
- Sustainable funding and infrastructure.

> **Challenges**
>
> Resilience strategies have to balance the benefits of a centralized federal data system with the kind of backup and cross-validation that a more decentralized system can provide. Some degree of data duplication can improve resilience, but too much can be wasteful. Developing backup data sources will require funding, coordination, and prioritization.

Efficient

A national data ecosystem should be efficient as well as resilient. It may need to include multiple data sources to provide different perspectives on critical issues, but it should eliminate data collections that are truly unnecessary or redundant.

> **Challenges**
>
> Data programs should not be discontinued without thoroughly understanding the consequences. There need to be guidelines and a process, involving data users as well as providers, to determine whether a data source can be sunsetted without losing critical information.

Sharable

The ecosystem should facilitate data-sharing on many levels. This includes sharing data among federal agencies, among state and local governments, and between the federal government and the states.

> **Challenges**
>
> There is currently no standardized way for agencies to handle data-sharing agreements (DSAs). Without a common set of legal frameworks, DSAs need to be developed on an agency-by-agency basis.

User experience

An ideal national data ecosystem should be:

Easy to use — by both humans and AI

National data should be easy for anyone to access and use. As some participants noted, data users should be supported as if they were customers rather than burdened by complex and inefficient systems. The system should serve both people who want to access and analyze the data directly and those who want to use it through AI agents and tools.

> **Challenges**
>
> Legacy systems and under-resourced agencies are limiting progress. Modernizing those systems will require investment, capacity, interoperability, and simplification.

Easy to contribute to

It should also be easy for current and future data producers to contribute to the ecosystem. That means simplifying data submission processes, reducing administrative and technical burdens, ensuring that systems are intuitive and efficient, and generally updating and rebuilding the government's current systems for data management.

> **Challenges**
>
> Here, too, legacy systems, lack of resources, and the need for modernization are limiting progress. Trust-building qualities such as strong metadata, privacy protections, and auditability must be maintained.

Relevant

A national data ecosystem should prioritize data that will have the widest use and the greatest impact. That includes local data available to serve state, city, and community needs, as well as supporting national indicators. Roundtable participants stressed the need to begin with the critical questions that data users need to address, and to be sure that the data is organized and made available to answer those questions.

> **Challenges**
>
> Many federal data collections cannot easily be applied to meet local needs. In addition, it's hard for the public to navigate the way data is distributed among federal agencies: A single question may need to be addressed by data from several parts of the government.

Timely

Much of today's federal data is not timely enough for many applications. It can take two or three years from the time the federal government collects data, for example, on education or Medicaid services, until the data is publicly available. Even some data that is available much more rapidly, such as monthly data on inflation, may still not be timely enough to track rapidly changing factors accurately.

> **Challenges**
>
> Federal data takes a long time to release, largely because it goes through careful review to ensure accuracy and validity. Third parties may be able to analyze the raw data more quickly, particularly with the help of AI. However, that may entail a tradeoff between accuracy, speed, and trust.

Public stewardship and social benefit

An ideal national data ecosystem should be:

Trustworthy

Trust needs to be built in at every step of the data pipeline, including data collection, storage, analysis, and sharing. The public needs to trust that government data will be independent, credible, accurate, complete, fair, representative of all demographics, auditable, and relevant to their needs.

> **Challenges**
>
> To earn trust, the national data ecosystem will need to be more accountable to the public, with clear ownership and oversight of data systems and transparency in how data is collected, managed, and shared. These foundational elements of accountability and trust need to persist through the expanding ways the public accesses data, ranging from direct access to AI-provided answers and analysis.

Disaggregated and representative

The national data ecosystem should include data that represents rural, low-income, and other communities that are often underrepresented and underserved.

> **Challenges**
>
> It can be difficult to gather survey data or other data from underserved communities due to the difficulty of reaching some households, distrust of government data collection, or other factors. In addition, changing government policies may make it more difficult to collect data broken down by race, ethnicity, and other variables.

Privacy-protected and secure

Individuals and communities that contribute data, whether through large-scale surveys or individualized programs like clinical trials, should have the assurance that their data will not be used in ways to which they have not consented.

> **Challenges**
>
> There is a need for more clear, enforceable, and comprehensive privacy laws and technical standards that can be applied across the U.S. Even where laws do exist, as one Roundtable participant put it, “there are many cracks in the interpretation of the law.”

Sustainable

The federal data system needs to be funded and maintained in a sustainable way with support guaranteed over time. The system needs stronger federal guarantees of reliable, sustainable support, alternate nonfederal funding sources for collecting key data, or both.

> **Challenges**
>
> With constant pressure to reduce the federal deficit, funding and infrastructure for data programs can be vulnerable to federal budget cuts. Rapid changes in funding can make it difficult to maintain data collections that are needed for longitudinal analysis.

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

The federal government's essential data roles

The federal government's essential data roles

The federal system sets the context for all national data. Roundtable participants generally believed that the federal government should ***recognize data as a national asset and a piece of public infrastructure that demands long-term investment***. They noted that the federal government is simultaneously a producer of information, a regulator of the data ecosystem, and an educator about the use of its own data. Roundtable participants identified the following critical roles for the federal government, which can serve as a starting point for further work.

> **There are four places for national data where the role of the federal government is irreplaceable: promoting stability, expanding coverage, addressing manipulation, and setting standards.**
>
> Beth Jarosz, Vice President, Association of Public Data Users

Funding expensive data collection infrastructure

The federal government funds high-cost enterprises, such as earth observation satellites and the national Census, that could not easily be replicated by philanthropies or the private sector. Significantly, the federal government does not try to profit from its data but collects and publishes it to inform anything ranging from business loans, to where a family decides to purchase a home, to major public policy decisions. That can make federal data sources more trustworthy than those from the private sector.

Helping to improve stability, consistency, uniformity, and accuracy of data systems

States have very different data systems, including different data collection methods, levels of technical and data capacity, and even definitions of the same statistics. In contrast, nationwide federal surveys like the U.S. Census use a single, consistent methodology for all 50 states. The federal government can also aggregate and harmonize data collected at the state level.

Setting data standards

Federal participation in advancing public-private standards can ensure data consistency and comparability for data collected at a state or local level, as well as data collected by the federal government itself. Setting and implementing data standards is a complex, multistakeholder task, and the federal government is well positioned to play a leadership role.

Providing national coverage at a granular level

Federal data can be geographically comprehensive through national surveys that provide consistent data across all geographies regardless of their size or resources. Federal data frequently covers the entire country, while also providing enough geographical granularity to make data useful at the state and local level.

Enabling international collaboration

Some data collections, like earth observation data, go beyond U.S. borders. The federal government collects and manages much of this data and can foster and manage international collaboration for these data sources.

Mandating essential data collections

Federal mandates have been important in directing states and local governments to collect data on health, housing, employment, and other topics that they might not otherwise study.

Evaluating federal programs

The federal government is able to collect performance data on its own programs, measuring success and identifying necessary improvements.

Supporting state and local data initiatives

In addition to mandating data collection, the federal government supports state and local data through funding, standard-setting, data aggregation and analysis, and other means.

Together, these collective observations on ideal qualities for a national data ecosystem and the unique role of the federal government set the context for future work. The following seven goals, and the priorities and opportunities associated with each, synthesize the most salient observations from Roundtable participants and our other research. We hope they can be a starting point for collaborative action.

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

GOAL 1

Maintaining critical federal data collections

Goal 1: Maintaining critical federal data collections

Any plan to reform America's data ecosystem has to begin by protecting the value of what we already have. Part of that effort involves demonstrating the importance of federal data as a national resource, as some new projects are doing now. America's Essential Data, for example, is collecting use cases for federal data, while the Data Foundation is working to develop a People's Data 100 list of America's most valued datasets.[13]

Beyond improving public understanding, however, recent changes to federal data have demanded an active response. The highest priorities to date have been to archive the data and tools that already exist, to monitor risks to current data collections, and to take action to preserve important data sources where needed. Future work can include new efforts to define which data sources should be considered "essential" for what purposes, applying AI and other technology to analyze data use, and finding appropriate ways to reduce unnecessary or redundant data programs without losing critically important information.

> **As a country, we believe in pushing decisions down to the lowest possible level – households, businesses, and local communities. But if people don't have good evidence, they're more likely to make bad decisions, so you don't get the best possible outcome for them and for society as a whole. The statistical system is the single most important source of information that people need to make good policy and economic decisions in the public and private sectors.**
>
> Erica Groshen, former Head of the Bureau of Labor Statistics

Ongoing priorities

Monitor data risks, engage the public, and archive data

Several independent organizations are now monitoring changes to federal data collections both across government and in specific areas. The organization dataindex.us[14] has developed a systematic monitoring system using required government information filings. Their Data Checkup methodology includes assessments of risks to data availability, data quality, and policies affecting the data.

With the risk of losing public access to some government data, websites, and tools, a number of existing and new organizations have started to archive this information. The Data Rescue Project is helping to support these efforts and aggregating information on data archiving initiatives with the help of hundreds of volunteers.[15] A group of Harvard researchers has created a copy of data.gov, the federal portal for all federal data collections. Public Environmental Data Partners (PEDP) rapidly recreated the Environmental Protection Agency's (EPA) EJScreen and the Climate and Economic Justice Screening Tool (CEJST) after they were taken down.[16] These new efforts complement archival systems that have been in place for years, such as the Internet Archive, which has more than four hundred million .gov web pages archived and available through its Wayback Machine.[17]

This archival work is critical, not only to provide independent backup systems for federal data, but to serve as models for similar federal programs. The federal government could learn technologies and methods for developing more effective archiving systems and processes from these independent groups. More robust, accessible data archives can serve both as historical records and benchmarks for any changes to important data collections.

Develop a taxonomy of data types and analyze the issues for each

To take a fresh, thorough approach to the nation's data ecosystem, we need better ways to analyze and describe the data that's currently available and what might be available in the future. Current and future national data sources include:

- Statistical data collected and analyzed by federal statistical agencies
- Administrative, programmatic, and regulatory data
- Data collected directly by federal agencies, which can range from survey data to earth observation data
- Data collected by cities and states, but aggregated by the federal government
- Scientific data from academic and other studies funded by the federal government
- Survey and other data collected by independent organizations
- Private-sector data
- Data crowdsourced from individuals or communities
- Data from social media, cell phone GPS data, and other sources based on use of technology

Each of these data sources, and others yet to be developed, have their own strengths and weaknesses, their own issues, and their own potential for use and misuse. The new, innovative Federal Data Field Guide provides a thorough catalog of federal data types.[18] This can be the basis for further analyzing both federal and non-federal data, and the pros and cons of different data types, to help guide the development of an improved national data ecosystem.

Emerging opportunities

Identify "essential" federal data sector by sector

With hundreds of thousands of open federal datasets listed on data.gov, and limited federal resources to review or analyze them, there is a growing need to understand the value of government datasets and their use cases.[19]

One efficient and effective method is to study critical data needs sector by sector. CODE is currently conducting a proof of concept for this approach in the health data space through a process driven by input from public health experts. This project, funded by the Robert Wood Johnson Foundation (RWJF), identified a "core canon" of more than 70 public health data sources.[20] CODE is now leading a working group to refine the list and analyze these core datasets and their use in detail. Such an analysis can include use cases for the data, risks to these data collections, possible additional data sources, and other factors, and will inform a methodology for further work in other sectors.

Track data use and apply AI to prioritize data sources

In addition to expert-driven analyses, new technical methods can shed light on how many people are using public data sources in what ways, providing a direct way to measure their public importance. Some new federal programs, notably the National Secure Data Service (NSDS) described below, are developing methods for assessing data usage while protecting privacy. Several Roundtable participants suggested tagging federal datasets with digital object identifiers (DOIs) to make it easier to track their use.

AI can also be a powerful new tool to identify data use cases that are of broad interest to the American public. By analyzing search engine use patterns, social media, or other sources, we can identify many of the top questions that Americans are asking, determine what national data can help answer those questions, and find gaps where new data is needed.

Develop a process for improving federal data efficiency

Making federal data more efficient is a valid goal, but it requires a systematic, programmatic approach. A nonpartisan, rigorous, and unbiased review of government data collections, with ample opportunity for public input, could identify data collections that are duplicative, those that may no longer be necessary for the goals they were meant to achieve, and opportunities to streamline and reduce the costs of data collection, production, and dissemination processes. This effort could build on OMB's existing standard process for reviewing data collections. Participants noted that this process should include not only technical experts, but subject matter experts as well as librarians, archivists, and others who have expertise in records review and retention.

As part of a streamlining process, experts should assess how data sources may have second- or third-order benefits that emerge unpredictably or organically. One panelist noted, for example, that the USGS North American bat monitoring survey — an obscure-seeming data collection — helps the government protect bats, which help America's farmers by eating harmful insects, reducing the need for pesticides and lowering farming costs.[21]

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

GOAL 2

Strong, reliable federal data governance

Goal 2: Strong, reliable federal data governance

The last decade and a half has seen major legislative, executive, and administrative initiatives to ensure the quality of government records and data. These new laws have built on a long history of data protection legislation that has consistently had strong bipartisan support. (See Appendix 5, "A Timeline of Data Legislation and Policy.")

The Evidence Act of 2019, signed by President Donald Trump during his first term, established the current framework for making federal data open, accessible, and usable. It added new protections for the independence of federal statistical agencies and required agencies to publish data in machine-readable formats, develop data inventories, and designate Chief Data Officers (CDOs). It also established a CDO Council within OMB.[22]

Data legislation and policy, however, are only effective if they are followed. Eight years after being signed into law, the Evidence Act has not yet been fully implemented. In addition, several lawsuits have alleged that recent government actions have violated the Privacy Act.[23]

In addition, several experts interviewed by CODE raised concerns that the provisions of the Evidence Act and other current laws may not be sufficient even if they are followed properly. They believe that current laws do not guarantee sufficient independence for statistical agencies, data privacy and security, public trust, capacity, and other core aspects of the American data ecosystem.

Ongoing priorities

Fully implement the Evidence Act

The Evidence Act requires that federal agencies make their data "open by default," meaning that they have an obligation to publish it unless there are compelling reasons not to, such as privacy or national security. OMB has laid out a four-phase implementation plan for the Evidence Act — including guidance on the open data that was issued in the last week of the Biden administration — but it has not been entirely fulfilled.[24] The Data Foundation is tracking implementation of OMB's four-phase plan as well as other activities related to the Evidence Act in its Evidence Act Hub.[25]

Establish and follow clear policies for privacy protection

The Roundtables highlighted the importance of data privacy protection, which has become a fundamental issue for data-collecting agencies and the public alike. There is no overarching privacy policy or set of policies governing the use of federal data. Different agencies use different procedures, terminology, and legal wording. State data systems, many of which feed into the federal data ecosystem, also have differing privacy laws and definitions.

The public has been worried about data privacy for decades, and some recent surveys show that their worries have become widespread. A 2023 Pew Research study showed that more than two-thirds of Americans are confused about how the government uses their data and are concerned about it.[26] More recently, in March 2026, a report from the Center for Democracy and Technology found that "An overwhelming majority of Americans are concerned about the privacy and security of their personal data that is held by the government."

Privacy is not only an issue of individual rights: It's also essential to maintaining data quality. If Americans can't trust that their data will be well protected, they will become even less likely to respond to government surveys than they have already become.

Privacy protections are especially important as the government moves toward greater data sharing between agencies. Under the Privacy Act, government data on individuals can only be used in ways that the government agency described when it was collected. Sharing data without individuals' consent for broad new purposes, such as government surveillance of individuals, could violate that basic principle.

While data sharing can lead to important new insights, it must be done through a clear and public process. Any program to link individual records must be conducted with effective privacy protections, meaningful nonpartisan oversight, and public accountability. Agencies need to develop memoranda of understanding (MOUs) or data-sharing agreements (DSAs) in order to share data with each other.[27] By developing model DSAs, individual agencies or coordinating offices, like the Statistical Policy Branch of OMB's Office of Information and Regulatory Affairs (OIRA), could both speed up data sharing and appropriately protect individual privacy, consent, and transparency.

If model DSAs are developed, they can be published for public input and comment.

> **There are three kinds of areas where the government can address privacy. The first stems from clarifying that certain types of data can only be used for the purpose for which they were collected such as Census or IRS Data. Second, we need to address high-risk areas like law enforcement and regulating the use of AI in those contexts. Third, we need to have more conversations about the government using data brokers.**
>
> Jennifer Huddleston, The Cato Institute

Improve access to federal data products

Both federal agencies and non-governmental organizations are working to make federal data more accessible and usable. Federal websites and data products do not consistently provide user experiences that support finding, understanding, and using data, and even the expert users at these Roundtables complained that it is often difficult to find critical data. USAFacts is dedicated to providing federal information in engaging, understandable ways for public access and use. USAFacts and the Partnership for Public Service also run the Federal Data Excellence program to "develop standards for evaluating the quality of data products from a user's perspective, a community for public data stewards to share insights and challenges, and a recognition program to celebrate user-friendly government data."[28]

Government agencies and offices have an opportunity to improve their own websites and develop AI agents to make their data more easily usable. The government's central resource for all federal open data, data.gov, has now launched a new catalog site and is inviting public input.[29] Individual agencies or coordinating entities like OMB could also bring in more outside data users for extensive user testing.

Emerging opportunities

Lay out a legislative agenda

In May 2025, Congressman William Timmons (R-SC) introduced a resolution with bipartisan co-sponsors to convene a new Commission on Evidence-Based Policymaking, like the one that originally shaped the Evidence Act.[30] Whether or not this proposal moves forward, government and civil society can use this time of technological advancement and government modernization to focus Congressional attention on new data reforms. For example, Congress could:

- Amend the Evidence Act to strengthen the protections it provides — drawing on the same kind of bipartisan collaboration that created the Evidence Act itself.
- Identify, enforce, and strengthen other existing laws that prohibit changes to major data collections without Congressional approval.
- Write more precise specifications for the most important, highest-priority datasets.
- Update national archiving requirements for the digital age.
- Create clear guidelines for making government data AI-ready.
- Pass a resolution or legislation recognizing the need to maintain accurate, scientific public data.
- Take other measures to maintain and improve national data resources.

Review the authorizing language for data collections

Most statutes set out general requirements for agencies to collect certain kinds of data without specifying precisely how the data should be collected. This makes sense, since legislators are not subject matter or data experts, and since data needs, including collection, analytical, and storage and retrieval methods, may change over time. However, such general language can leave latitude for new political appointees to change or discontinue longstanding data collections at will. Such disruptions to important data systems could also make it harder for Congress itself to use data for decision-making.

A review of authorizing language and statutory mandates for the most widely used datasets, and recent or planned changes to them, could help define the right balance between more prescriptive and more open authorizing language. This review would be particularly timely given the Supreme Court's recent decision on "Chevron deference," which limits federal agency authority to interpret ambiguous statutory language and shifts more authority to the courts and Congress.[31]

Restructure federal data systems for efficacy and efficiency

Since the federal data ecosystem has grown piecemeal over decades, it's not surprising that it is inefficiently organized and poorly coordinated in many ways. New ideas and structures could make the ecosystem more rational and efficient.

A prominent example is the opportunity to restructure elements of the federal statistical system. The United States has a decentralized statistical system in which individual agencies collect and manage statistics on specific topics.[32] Proposals to combine certain aspects of this system have been discussed for years and have broad support.[33] For example, the detailed version of the President's 2026 Budget proposal included a plan to consolidate the U.S. Census Bureau, the Bureau of Economic Analysis (BEA), and the Bureau of Labor Statistics (BLS) within the Department of Commerce.[34] However, that proposal is not explicitly repeated in the President's 2027 Budget, so it is unclear where it stands.[35]

Many leaders in the statistical community would welcome reform if it is done with care.[36] The executive director of the American Statistical Association (ASA) has emphasized the need to "use the 'scalpel,' not the 'hatchet'" in modernizing statistical agencies. Such a careful approach would include testing new data collection methods, ensuring ongoing funding for statistical programs, protecting agencies from political interference, and taking a transparent, iterative approach overall. One group of former federal statistical leaders has launched a project along these lines, Modernizing the Federal Statistical System, which is in the process of publishing its recommendations.[37]

Establish more consistent funding for predictability and sustainability

Statistical and other data-gathering agencies have long faced budgetary constraints that undermine their ability to collect data, update their technologies, and modernize their methodologies. For example, the BLS budget was actually nine percent lower in 2024 than it was in 2010, adjusted for inflation, even while being asked to do the same work it did 15 years ago.[38] Without a concurrent and intentional effort to make data agencies more efficient, underinvestment can lead to diminished ability for an agency to consistently provide data to the public. There is a need for policies and processes to ensure that ongoing data collections have sustainable, reliable funding from year to year and administration to administration. This is critical to:

- Maintain current data collections.
- Guarantee the ability to track long-term trends with a consistent methodology.
- Support modernization of legacy systems and technical infrastructure.
- Enable continuous improvement of technologies and methodologies.
- Support the development of new research methods. (Any new research methodology should be benchmarked against the methods it is meant to replace — a process that requires running two surveys in parallel, for example, which will temporarily increase expense.)

Ensuring adequate, predictable funding will require shifting to a new process for supporting statistical agencies, such as multiyear funding models, rather than having these agencies depend on funding that is reevaluated year to year. It will also require building support in Congress for these goals. The American Statistical Association and organizations such as the Census Project and the Friends of BLS are advocating for better and more reliable funding for critical agencies.[39] They have an opportunity to develop broader support from the public and other stakeholders, particularly as financial leaders are raising concerns about core economic data. Former Federal Reserve Chair Jerome Powell, for example, has estimated that federal statistics may overestimate job growth by up to sixty thousand jobs a month.[40]

Build out the National Secure Data Service (NSDS)

In addition to the privacy policies described above, privacy protections should remain foundational in the technology for managing federal data. The private sector has developed a number of privacy-enhancing technologies that can also be used to manage personal government data responsibly. Sensitive data can also be managed using tiered access systems, limiting certain kinds of data to researchers and other qualified and vetted users. This can be particularly important for data on small demographic groups, such as indigenous communities, where the risk of reidentification or cultural sensitivities are high.

The National Secure Data Service (NSDS) has the potential to combine policy and technology for a more private and secure data system. The National Center for Science and Engineering Statistics (NCSES) of the National Science Foundation (NSF) is now building out a demonstration of the NSDS, which was required under the 2022 CHIPS and Science Act. As the NCSES website states, the goal of the project is "to inform efforts for developing a shared services model that would streamline and innovate data sharing and linking to enable decision-making at all levels of government and in all sectors."[41]

The NSDS is currently funded only through the fall of 2026. Many Roundtable participants supported building out the NSDS, but expressed concern about its future and whether it will have adequate funding and staffing. The NSDS as a demonstration project has thus far had a low public profile, and those who do know about it may think of it only as a technical project. There is an opportunity to broaden public awareness for the project and increase participation and feedback for its initial build-out.

Use new technology for better data access

Federal agencies have embraced the use of Application Programming Interfaces (APIs) to make it easier to access and apply public data. That trend supports the use of government data for AI through Retrieval-Augmented Generation (RAG) systems, which make it possible to develop the large language models (LLMs) that power AI tools more quickly.

A newer development, the Model Context Protocol (MCP), may make government data accessible for AI at an even greater level. As one participant put it, “APIs revolutionized the world in terms of how data goes back and forth. Now MCP servers have revolutionized the way agents and humans interact with information.” Many participants were enthusiastic about the potential for applying MCP, calling it a “game-changer.“ Even if MCP gives way to a new approach, focusing on agent-to-agent data communication will be a must for government data publishers as the space evolves.

The Census Bureau is now experimenting with an MCP server for its data, and some Roundtable participants suggested that every federal agency should consider building MCP servers to enable the public to ask questions of any government source using natural language. Whether MCP servers become universally accepted or are replaced by other technologies, government agencies should continue to use best-practice technology for public data access and use.

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

GOAL 3

A national data ecosystem that serves all Americans

Goal 3: A national data ecosystem that serves all Americans

A national data ecosystem should serve all Americans. It is built with their tax dollars, represents them, and is used to make decisions on their behalf and for their benefit. Disaggregated data — data that is broken down by race and ethnicity, income and education levels, gender identity, sexual orientation, disability, or other factors — is essential for understanding how different groups are impacted by different issues. For example, such data can be analyzed to understand racial and ethnic disparities in healthcare, issues impacting the gay community, issues faced by Americans in poverty, and the concerns of rural Americans. Importantly, disaggregated data also helps understand key issues facing majority groups too, like the non-Hispanic white population or native-born Americans. Unfortunately, for many years and across administrations, America's underserved communities have often been underrepresented in the nation's data. The current administration has continued implementing new federal guidelines designed to improve the collection of race and ethnicity data, while, in contrast, removing data on sexual orientation and gender identity from many federal data sources.

Ongoing priorities

Maintain a federal commitment to improving disaggregated data

For more than two centuries, the U.S. Census has continually improved the precision with which it measures race and ethnicity in the population.[42] Most recently, in March 2024, OMB released revisions to Statistical Policy Directive 15 (SPD 15) that provide for more precise collection of race and ethnicity data across government. Data advocates are now watching to see how these revisions will be implemented. Federal agencies had originally faced a September 2025 deadline to develop action plans for their SPD 15 revisions, but OMB has now extended the deadline to March 2027.

The Leadership Conference Education Fund is monitoring the implementation of SPD 15 revisions and notes that 75 federal data collections are already implementing the revisions in whole or in part. In sum, they conclude that there are "promising signs of progress, but federal agencies and OMB must commit to fully implementing SPD 15's detailed data collection requirement."

Limit reductions of disaggregated data

In the same way that expanding disaggregated data is beneficial to understanding the 340 million Americans, reductions in disaggregated data can prove harmful. Recent government policies have led to a loss of disaggregated data on sexual orientation and gender identity (SOGI). In early 2026, two studies showed that SOGI data elements had been removed from some 360 government data collections in the course of a year.[43] A UCLA analysis has stressed the importance of maintaining this data in a wide variety of federal surveys — on topics including health, nutrition, housing, and criminal justice — and described the impact of losing that data in the future. Public health, civil rights, and other organizations are using advocacy and legal action to try to reverse these changes, out of concern that lost data could make it more difficult to address issues including homelessness, discrimination, health and disease.

Emerging opportunities

Support disaggregated data at the state and local level

States and cities need accurate, detailed data about their populations, which includes data on race and ethnicity as well as factors like education and income. Several have developed programs for collecting disaggregated data, often with the guidance of The Leadership Conference Education Fund, which has established the Data Disaggregation Action Network to preserve this kind of data.[44]

Through education and activation of partner community-based organizations, the Fund is encouraging its state partners to implement the revised SPD 15 guidelines at the same time that federal agencies are implementing them. In the July 2025 report "Disaggregation Nation," the Fund describes how race and ethnicity data has been improved in four "case study states": California, Illinois, New York, and Oregon.[45] The states' goals include a more complete picture of mortality rates for different groups, helping public health and health service agencies coordinate and improve their work, and giving community organizations a deeper understanding of their communities' needs.

Engage with local communities

A national data ecosystem that serves all Americans requires meaningful community engagement. Engaging with local communities around data issues can give them more control over the data; empower them to address community-specific issues and needs; better tailor services and support for greater use and impact; and ensure dignity and agency for community members. Particularly around politically sensitive issues like environmental justice or public safety, data programs should follow the rule of "nothing about us without us": respecting the principle that communities should control how their data is used and who accesses it.[46]

Partnerships with community-based organizations (CBOs) provide an opportunity to collect community-level data through surveys or crowdsourcing. These partnerships can help gather data about a CBO's community and its individuals' needs. In one well-known environmental project, for example, NOAA has worked with communities across the country to measure heat in cities and document the risk of urban heat islands in poor neighborhoods. Developing partnerships like this requires trusted relationships and a clear understanding about how data can benefit the community.

In addition to building trust, community partnerships can develop new and valuable data sources. Local communities provide natural opportunities for citizen science, which is often called participatory science. Participatory science can include collecting hyperlocal data with low-cost sensors for air and water quality monitoring, or data from drones analyzed with AI. Similar open-source hardware could unlock other local data, especially where federal tools are outdated, expensive, or slow. This creates a virtuous cycle that benefits the data ecosystem and the American public more broadly.

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

GOAL 4

An AI-enabled system for better data and data access

Goal 4: An AI-enabled system for better data and data access

The explosive development of AI is changing our society rapidly and is impacting the national data ecosystem as well. Much of the current interest centers on Generative AI (GenAI), which analyzes patterns in massive amounts of data to develop new data and information with human prompting.[47]

- AI applications require more complete and better structured data from government and other sources,
- AI can help improve data sources that already exist,
- AI can be used to create new data sources for public use, and
- AI will increase public demand for trusted data and information through new AI tools and agents.

Ultimately, AI will need to provide more accurate and useful answers to a range of human questions before experts will trust it for more serious use. While tools are available to evaluate and benchmark AI applications, and to help determine whether their results are accurate, safe, and fair, these tools are not yet fully developed or widely used. Government can play a role by ensuring that AI tools and their results are benchmarked, assessed, and evaluated transparently at every step of development. As one participant put it simply, "No benchmarks, no trust."

> **We're at a place where you can see a kind of flowering with these incredible tools that are available. There's a real hunger for people to understand what's going on around them. It's been a great time to see not only how people are using data, but also how they're using AI at the moment.**
>
> Simon Rogers, Google

Ongoing priorities

Address critical concerns about data and AI

While AI is a powerful tool, Roundtable participants stressed that it is not a panacea or shortcut for solving the nation's data challenges. It can only be effective if it is used in a system with high-quality data, governance rules, privacy protections, and human accountability. Any vision for a future national data ecosystem needs to address the potential and limitations of AI in that context. The Roundtables highlighted several areas to prioritize for development.

Governance

The development and use of AI have outpaced governance and legal systems. Effective governance should prioritize responsible, transparent, and human-centered use of data and AI with ethical, privacy, and safety guardrails. A good governance infrastructure should include better vendor contracts, legal frameworks, and data-sharing agreements, as well as other improvements. The workings of AI should be traceable and auditable, and AI tools should disclose the data they use and its provenance. As one participant said, "AI is only as strong as the systems underneath it. Reproducibility, clean documentation, and shared standards are the entry ticket, not optional upgrades."

AI literacy

Non-technical government workers and the public alike may not understand that current AI systems often give wrong answers. AI researchers are studying the "hallucination" problem and are still not sure whether it can be solved, at least for LLMs. AI users need to be educated about this issue — both so that they don't rely on AI without checking its results, and so their expectations don't exceed what AI can currently do.

Privacy

AI must embed best-practice privacy and security protections, such as differential privacy, secure access layers, and data minimization. Some researchers are advocating the development of "small language models" designed for specific purposes rather than large, generalized models, as part of a privacy strategy.

Social and ethical issues

The use of AI has raised a host of issues, including potential bias in AI algorithms, the environmental impacts of massive data centers, and the potential for a new kind of digital divide. These issues need to be addressed thoroughly, publicly, and fairly.

Trust

Concern about all these issues can undermine public trust in AI and could interfere with appropriate progress. Roundtable participants were concerned that AI is scaling too fast without having dealt with core issues, and before AI has demonstrated enough tangible benefits. They noted that AI may be implemented "for the sake of novelty, not need." But as one participant observed, "Without training, standards, or clear use cases, the hype will lead to disappointment rather than transformation."

Make government data AI-ready

AI can only be as good as the data it is trained on or can access programmatically, and much federal data is not yet well-suited to that task. A major issue is the lack of sufficient and adequate metadata. While AI and LLMs can help improve government data in many ways, including metadata, the metadata problem begins with data collection and needs systemic solutions with widely accepted programmatic benchmarks and human-in-the-loop review. As one participant put it, "AI is not a get-out-of-jail-free card for bad data management."

The March 2026 White House National Policy Framework for Artificial Intelligence, which makes legislative recommendations to Congress, includes recommendations for making government data more AI-ready.[48] Specifically, it states that "Congress should provide resources to make federal datasets accessible to industry and academia in AI-ready formats for use in training AI models and systems."

A thorough review of historical, current, and planned future data collections could improve the collection of AI-ready metadata and government data overall as well as identify datasets that could benefit from AI-ready metadata improvements. In 2024, the U.S. Department of Commerce published a report on how to make government data more AI-ready.[49] USAFacts and CODE have each collaborated with the Department of Commerce on a project on this issue, and issued follow-up reports on their findings.[50,51] USAFacts has published a Guide to AI-Ready Data for Federal Agencies as well.

The federal government has also encouraged public-private partnerships around AI, through projects like the National Artificial Intelligence Research Resource (NAIRR) Pilot. The NAIRR Pilot is run by the NSF and engages numerous tech firms that contribute resources and expertise. With expert resources like these, AI itself can be used to improve government data and make it more usable for AI-enabled analysis. Examples, model protocols, and toolkits for government agencies could all help speed the use of AI for improving metadata and government data overall.

> **We may be experiencing a data winter at the time of an AI summer. If we don't fix this contraction in meaningful, equitable access to data for public interest use, the AI summer might also not be that long.**
>
> Stefaan Verhulst, The GovLab

Emerging opportunities

Use AI to harmonize data sources

With the future of some federal data in question, researchers in many fields are looking for ways to gain new, reliable insights without relying on government data as a single source of truth.

Researchers and other users will have to use multiple additional sources to supplement federal data. But data from different sources, such as state and academic data sources, may not always align. As one participant asked, "What happens when the University of Chicago and University of Michigan produce results that aren't the same?"

With AI, Roundtable participants believed that "data harmonization is a solvable problem." Generative AI could help automate dataset reconciliation across jurisdictions, for example, aligning education, workforce, or health data across states. AI could standardize formats, validate fields, and classify data types, but only if there is a clear framework and strong oversight. Humans need to validate outputs, and local actors need to be part of the loop.

Participants suggested launching one or more pilot projects to test out this approach. Researchers could use AI, for example, to help align one dataset type (e.g., student assessments or workforce credentials) across several states. Further work could leverage AI to harmonize and process data accurately across federal, local, and county levels as well. Unlike AI applications that meld and present diverse information in a narrative form, this approach would use AI to create new, accurate datasets based on comparing data from different sources.

Use AI to create new data sources

Data experts have become increasingly interested in nontraditional data sources (NTDs), many of which are created from data analyzed with the help of AI. The GovLab at New York University recently categorized many of these NTDs, such as internet search data, mobile data, and social media data, and gave examples of their emerging use.[52] These data sources can be used for healthcare improvement, humanitarian aid and disaster response, climate and environment, urban systems, economic and labor dynamics, and more.

Government agencies are experimenting with AI-enabled analysis of nontraditional data. The U.S. Census Bureau, for example, is trying new approaches as a way to counter the problem of declining survey response.[53] In addition to using administrative and business data, Census is using AI to analyze satellite imagery to verify business and residential addresses.

Data created as ordinary individuals use the internet, such as social media and search data, can also be analyzed and used. The Organization for Economic Cooperation and Development (OECD) has used Google search data to assess national perceptions of well-being in dozens of countries, while others have used it to assess U.S. GDP growth in real time.[54] Google has also applied the concept of nowcasting to weather data and is analyzing satellite data to provide real-time weather predictions in Africa to make up for a lack of sufficient radar observation stations.[55]

Finally, as an important strategy for privacy protection, AI can create synthetic data for analysis, for example, in health research that involves sensitive individual information. Synthetic datasets mimic original data, including the characteristics of individual records, while removing any actual personal information. However, synthetic datasets need methods for ensuring they are free of bias and representative of the underlying data. Transparency here is key.

Use AI to democratize data access and analysis

For more than a decade, the open data movement has advocated to "democratize data" by making government data sources more widely available and usable. While that effort resulted in key steps forward like the DATA Act of 2014, the Foundations for Evidence-based Policymaking Act of 2018, and the creation of the first Federal Data Strategy, it has been limited by the public's technical proficiency and by government's inconsistent investment in public data products across agencies. Moreover, government data products often continue to lack data visualizations or other content to help users understand the data. USAFacts was established to provide visualizations, narratives, and tools for communicating important insights from federal data to the general public. Government data agencies should strive to do the same.

The new generation of GenAI tools can make government data even more directly accessible for public use. It's becoming easier for anyone with a working knowledge of GenAI to do basic data analysis, create narrative descriptions of datasets and their meaning, create data visualizations, assess the relative merits of different data sources, or learn from public data in other ways. At the same time, agencies can continue to make their data more "AI-ready" for these applications, and can use AI themselves to create more accessible and usable data resources.

These technical advances provide new opportunities for collaborative efforts to help the public put data to use. Some government programs are already developing GenAI tools that will allow the public to query their data, such as the GenAI concierge being developed as part of the NSDS.[56] But this effort doesn't need to be limited to the government. GenAI companies, NGOs, interested users, subject matter experts, and other stakeholders can and should experiment with GenAI to learn the potential and the limits of current tools for democratizing data use.

> **AI has taken all the friction out of accessing data, but that doesn't mean it is easy to get good information out. We want to take all the context and learnings that we have about how to ask a good research question to get good data out, and turn that into an AI experience that is trustworthy and easy to interpret in context.**
>
> Kari D'Elia, USAFacts

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

GOAL 5

Extensive and usable systems of state and local data

Goal 5: Extensive and usable systems of state and local data

States and cities play an important role in America's national data ecosystem. Some national data collections, such as health and education statistics, rely on aggregating and analyzing data from all 50 states.[57] Cities can be the source of valuable data on transportation, pollution, public health, and other important issues. However, these subnational data collections have their own flaws and are not consistently produced or interoperable with federal data or other state and local data.

> **What we're seeing is that data is as precious a resource as electricity and water. The hope is that we can work together to really be able to demonstrate the value of the data and make sure that we're democratizing access to those data.**
>
> Jonathan Porat, Chief Technology Officer, State of California

Ongoing priorities

Ensure federal and other funding for state data collection

States and cities vary greatly in their ability to collect and maintain data resources and in their use of data. These disparities lead to unequal data quality and access, which in turn impacts national decision-making and runs the risk of perpetuating inequities. But regardless of capacity, all states and cities require resources to maintain their data collections and improve them where necessary.

State data systems often rely on a combination of state and federal funding. The federal government can:

Fund and incentivize core data collections

Many state data collection programs depend on federal funding — for example, funding from the Centers for Medicare and Medicaid Services (CMS) to support the collection of Medicaid data.

Fund innovation and share tools

Federal grants can support new collection methods and open-source tools that states can adopt. Models and code should be reusable, not reinvented for each jurisdiction.

Fund pilots and scale what works

First, identify existing working models, then fund and expand them across other states.

Federal funding cuts can damage the states' ability to collect important data. Federal agencies can improve state data by maintaining funding going forward and potentially restoring some of the previously cut funding. This is especially important for longitudinal collections that track public data over time.

A complementary approach to ensuring sustainability is to protect data systems in state law. For example, Maryland's longitudinal education data system has been protected by law and a funding source since 2010. Federal data collections that are important to states could also be safeguarded in state law. For example, states could enact legislation requiring the state to step in and collect Integrated Postsecondary Education Data System (IPEDS) data, should that data collection fail to be continued.[58]

Improve state data sharing and set data standards

In addition to helping each state improve its own data resources, data sharing can help prevent fraud, track healthcare expenditures, enforce child support when people move across state lines, provide benchmarks for performance, and more. States already exchange vital statistics (birth and death records) and driver's license data across their borders. This kind of data sharing can also help each state improve its own internal statistics, state programs, and citizen services.

Despite the benefits, capacity issues and inconsistent data systems with environmental, education, and police data, for example, have made it difficult to develop more sophisticated data-sharing agreements (DSAs). Legal barriers, data quality concerns, and security risks can also limit progress.

By collaborating on data collection strategies and standards, and developing new collaborations and data-sharing agreements, state governments can both improve their own data and make it easier to use for the common good. For example, Arkansas has developed a data-sharing platform that allows the state to track people who live in Missouri but get their medical care in Arkansas, or live in Arkansas but get their medical care in Oklahoma, so that they have a single Medicaid record that moves across state borders. The state then populates that record with other information from driver's licenses, unemployment insurance, or other state sources.[59] This ultimately eases the administrative burden on companies, individuals, and civil servants when providing data to the government.

In another example, the Wage Records Program of the BLS provides a voluntary way for states to safely share harmonized longitudinal workers' wage records.[60] Thirty-five states now participate. This sharing is used for analyzing issues such as interstate commuting and migration, and education and training outcomes.

As an immediate opportunity, states and organizations that have led successful data-sharing programs could publish their DSAs. A collection of public model DSAs could help facilitate new programs and simplify the legal processes involved.

Regional coalitions like those now being formed around public health data can address problems across an entire region as well as state by state. These new state alliances have formed to develop and share health information in response to changes in federal health policy and cuts in federal support.[61] For example, the Northeast Public Health Collaborative, a project developed with partnership from the Robert Wood Johnson Foundation, has been launched as a voluntary regional coalition of public health agencies and leaders, brought together "to share expertise, improve coordination, enhance capacity, strengthen regional readiness, and promote and protect evidence-based public health."[62] A similar new collaborative, the West Coast Health Alliance, represents Washington, Oregon, California, and Hawaii.[63]

While the federal government already provides standards for some state and local data, limited resources make it difficult for some states and municipalities to comply. The federal government can help lead national standardization efforts and support other efforts being led by independent organizations. It can also compile successful state examples, launch competitive grant programs, and contact governors about supporting data standardization legislation.

Emerging opportunities

Preserve and use data at the state and local level

Cities and states manage many kinds of data to meet federal requirements, but they often exercise surprisingly little control over the data they collect. In many sectors, data is collected by local governments, passed up to the state government, and then sent to the federal government for analysis.

In principle, this system should enable state governments to access data about their own and other states after it has been cleaned and analyzed by the federal government. For example, EDFacts, run by the U.S. Department of Education, collects data annually from state education agencies on enrollment, assessment scores, and student characteristics. The data is collected, reviewed, processed, standardized, and released back to the public.

In most cases, however, states see limited direct benefits from the federal use of their data. States may never get access to the processed and improved versions of their data that the federal government collects, and they may not even keep and manage their own data in usable formats. The federal versions of state data may be more suited to national comparisons than to local use. And even when federal data analysis does benefit the states, that analysis can take years to produce.

Several changes to this system could improve overall efficiency and bring greater benefits to states and cities:

States should keep the data they produce

As a basic, critical step, states should maintain copies of the data they send to the federal government so they can use it themselves or share it with others. Surprisingly, states often send federally mandated data to federal agencies without keeping the data for their own use. Processes for keeping data within the states, with appropriate privacy protections and security for sensitive data, should be guaranteed by law and immune to changes in administration.

Federal agencies should share improved data back with the states

As one participant put it, “States are forced to provide data to feds without seemingly getting anything tangible back. Data and information need to flow in both directions so state and local governments can get that same data back to actually use it to take action.”

States should consider also providing data to independent entities for analysis

The federal government is not the only entity that can aggregate and analyze state and local data. Universities, nonprofits, data collaboratives, or other organizations can often analyze data more quickly.

The COVID pandemic gave us a successful proof of concept for this kind of independent analysis. At a time when the Centers for Disease Control and Prevention (CDC) was not in a position to provide state-by-state statistics on the pandemic, both the Center for Government Excellence at Johns Hopkins University and USAFacts developed their own dashboards by aggregating data themselves.[64,65] The dashboards became go-to sources for timely, reliable COVID information. These and other examples, like a project on climate vulnerability from the Environmental Defense Fund, can serve as models for independent analysis of state and local data that can substitute for or improve on government analyses.[66]

Again, this is an area where AI-first infrastructure and technology could have a massive impact. MCP servers, APIs, and improved metadata are needed to make state and local data accessible and understandable by AI systems. Basing state data publishing on federal AI-readiness standards would lower the bar to entry, though more public-private partnerships would be required to make a truly interoperable data system across tens of thousands of government entities.

Use administrative data for new insights

Federal data stewards have been working to supplement survey-based statistical data with administrative data, which tracks government transactions and records such as pensions, taxation, and births and deaths.[67] This data is of high interest because government surveys, including the U.S. decennial Census, have increasingly suffered from the problem of low response rates. Many populations, including non-English speakers, members of minority groups, young children, and those who do not live in traditional housing, are difficult to reach and to count.[68] At the same time, the proliferation of surveys from polling firms and commercial interests has made many Americans reluctant to respond to surveys in general. Low survey response leads to lower quality data, which can lead to major revisions and reduce trust in federal data in turn.

Administrative data could help provide solutions for state as well as federal data. A recent opinion piece from NORC at the University of Chicago noted that "states are sitting on a treasure trove of administrative data, including eligibility records, service usage patterns, benefit payments, program enrollments, and outcome measurements."[69]

Administrative data is not a panacea: It may contain inaccuracies, may be collected using different definitions than other data collections, and may not be representative of the population. There are significant privacy challenges when government agencies share administrative data in ways it was not originally intended to be used. In addition, administrative sources cannot replace data that can only be gathered through surveys. For example, only surveys can gather data that reflects individuals' opinions and intentions, such as determining how many unemployed people are looking for work.

All that said, administrative data is a potential resource that state officials and data stewards can consider as they look to improve their own data resources. At the federal level, administrative data can contribute to the national data ecosystem as well.

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

GOAL 6

Robust American data developed with and for America's businesses

Goal 6: Robust American data developed with and for America's businesses

The U.S. has many different kinds of business communities, ranging from small businesses to large conglomerates that operate in all different sectors. All of them may have a stake in using national data, and may play important roles in protecting and improving the data ecosystem. For example:

- **The financial sector** relies on consistent, accurate economic data for its analysis, projections, and trillions of dollars in investment decisions.
- **Insurance and reinsurance companies** need accurate, extensive climate, health, and demographic data for risk assessment and projections.
- **The healthcare industry** needs a wide range of data on disease risks and trends, treatment effectiveness, and the social determinants of health.
- **The housing industry** relies on demographic, education, financial, and other data for development plans, real estate, and business projections.
- **Geospatial data companies** rely on massive amounts of Earth observation data to create their products and services.
- **Marine industries** (such as fisheries and seaports) rely on NOAA data.
- **The restaurant industry** needs data to evaluate food prices, food safety, and labor market and wage trends.
- **Farmers and agribusiness** need accurate climate data, weather data, and data on soil quality.
- **Tech companies developing AI applications** require large, accurate public datasets to train their models.
- **Consumer and retail businesses** use weather data for operational and supply chain management, demographic data for marketing, economic data for strategic planning, and other federal data specific to their business goals.

The first Trump Administration promoted the business value of open government data: In 2017 and 2018, CODE collaborated with OMB to hold two Roundtables, one on Open Data for Economic Growth and one on Leveraging Data as a Strategic Asset.[70,71] The convenings and the reports they produced underscored that administration's commitment to making federal data a valuable resource for American businesses large and small.

> **As we seek to expand the federal data ecosystem, I would love to see more data on small businesses. At the end of the day, small businesses guide how a community responds to any type of changes, including national policy, employment trends, and environmental changes. During disasters, small businesses hold communities up and it's those stories that help us understand on the ground what people need.**
>
> Afia Tyus, Southern Economic Advancement Project (SEAP)

Ongoing priorities

Explore the public use of private-sector data

A better national data ecosystem will rely on government data and commercial data complementing each other. In many situations, the government is the only possible collector and provider of data, such as surveys that require high levels of funding, comprehensive geographical reach, or standardization across geographies and demographics. Similarly, commercial data can also provide unique public value. For example, large corporations collect data on their customers that can provide new insights if it is managed in ways that respect individual privacy. Businesses collect data on real estate and land use, medical data, scientific data, and other kinds of data that frequently fill gaps in government data and could be applied for public good. And nonprofit organizations may collect data or develop digital tools to serve their particular missions. Taken together, government and private-sector data can provide a more comprehensive view of the country.

At the same time, data advocates have cautioned that private sector data should not be expected to replace everything that the federal government can do.[72]

The limitations of private sector data include:

Ethical and legal considerations

Privately collected data should not be used if there is any risk of breaching privacy or exposing personally identifiable data.

Incentives

Private companies generally don't have clear incentives to share their data, and may have an incentive to keep it only for their own use.

Scope

Private data collections can't match the scope of federal surveys. For example, the American Community Survey, one of the most widely used federal data resources, is conducted by the U.S. Census Bureau in ways that the private sector may not be able to replicate.[73,74]

Sustainability

In a worst-case scenario, private companies may go out of business and take their data down with them.

Trustworthiness

Private-sector data is collected to meet business needs first, not public-interest priorities. Companies may bias the data in ways that serve their interests — and even if they don't, the public may suspect that they are. Without transparency, review, and some form of governance, private-sector data may not be trustworthy enough to be very valuable.

Despite these limitations, with the current disruptions to federal data sources, it's important to look at the potential for the private sector to supplement government data collections and provide new, valuable data and tools. In addition to data that businesses naturally produce, private research firms can also provide data for the national data ecosystem, especially as a temporary stopgap in moments of disruption to federal data.

The combination of public and private data can be more powerful, and more complete, than either type of data used alone. For example, private data sources may have more timely and granular data than government data collections, though they also require careful stewardship to be used responsibly.

Ongoing efforts will be needed to determine what kinds of private-sector data and analyses can supplement federal collections, and what their benefits and limitations are. This can cover a wide range of private-sector data, including data analyzed by financial institutions for their clients, data from healthcare records, retail data, privately conducted surveys, and many others.

> **Much of the process of getting private sector data is part of a broader coalition-building exercise. It's about bringing in the business community — they recognize the value of these datasets and use this information to make their business decisions. In many cases, they could provide data in a way that is not going to cost them a lot.**
>
> Chris Jackson, SSRS

Identify high-value government data for American businesses

In a recent opinion piece, Michael R. Bloomberg made the business case for government data and accurate federal statistics. "Business leaders," he wrote, "are...dependent on this data as they make planning and investment decisions, from retailers figuring out where to locate a store, to farmers and ranchers weighing how much of their production to hedge, to manufacturers deliberating whether to expand their plants."[75]

Given the value of government data, business groups are in a good position to advocate for the data they need. These range from the largest trade associations like the U.S. Chamber of Commerce to small groups of companies with common interests. In addition to individual businesses, trade associations and other groups of businesses in specific sectors can advocate collectively for the data they need. In general, companies may hesitate to advocate for an issue on their own, but may be willing to join a larger group of businesses advocating together.

Emerging opportunities

Conduct pilots for using proxy or blended data from business sources

Federal agencies and researchers are increasingly interested in "proxy data," which in this case means indirect measurements that can shed light on public issues. One well-known example is the ADP employment report. Others can help track the impacts of natural disasters. For example, the Waffle House Index reflects disasters' impact based on drops in the number of customers to that chain, while GasBuddy's Fuel Availability Tracker shows where and how disasters interfere with the fuel supply.[76]

Proxy data sources vary in quality, transparency, continuity, and reliability, and even the best may have limitations. The ADP report, for example, provides timely metrics on jobs added or lost in a month, broken down by industry. It became an important source of information during the fall 2025 government shutdown, when federal jobs statistics for the month of October were not available. However, since the ADP report is based only on the roughly 20 million workers whose employers use ADP, it does not provide a comprehensive or unbiased view of the labor market.

One possible approach, now being explored in different contexts, is to use "blended data" in a way that combines both traditional statistics and private-sector data that can contribute additional insights and often more timely information. The Bureau of Labor Statistics, for example, is exploring ways to use consumer prices from the internet as a data source. In public health, the Yale project PopHIVE combines many kinds of data from both private sector and government sources and synthesizes them to make it possible to use the data more efficiently.[77] For example, PopHIVE tracks infectious disease patterns by combining CDC data with data collected from medical records, Google search patterns, a network of smart thermometers, and other sources.

Several Roundtable participants suggested that state and local projects could be a good first step to developing public-private collaborations. Housing data emerged as a particularly promising area for one or more pilots. Data on basic issues such as building vacancies, illegal units, building permits, rental units, affordability, and evictions is inconsistent and difficult to aggregate and analyze. Pilot projects on housing data could both analyze existing data and develop new public data sources, using public-private partnerships to build a multi-source housing data framework. Partners could include academic centers, state governments, and real estate data companies and private-sector mapping and tech companies. Zillow, for example, has extensive data on housing that could be powerfully combined with federal data.

Explore cooperative funding for data collection, management, and publication

With concerns about the reliability of government data and the high costs of private datasets, some businesses that rely on data are working together to develop the data themselves. For example, Amazon, Microsoft, Meta, and TomTom have co-funded the Overture Maps Foundation to reduce their dependence on Google Maps data and produce a new data source that they own collaboratively.[78] Similar models could be applied to financial data, sustainability data, and other types of data that serve important business needs.

One way to think about this option is as a "utility cooperative" for developing, managing, and sharing data.[79] Utility cooperatives (co-ops) produce products that serve the members of the cooperative, who simultaneously fund, own, govern, and use those products. This model was used to electrify most of rural America during the New Deal. Utility cooperatives are independent of government institutions and should be protected against political changes that can impact data collections and programs.

A pilot project could develop a utility co-op for data in an area where businesses have a particular data need that is not being met. Several corporations could collaborate to fund an experiment to develop vetted, reliable, and standardized sustainability data through a utility co-op without relying on government agencies to do so.

BUILDING THE NATIONAL DATA ECOSYSTEM AMERICA DESERVES

GOAL 7

Shared expertise, communication, and collaboration around national data

Goal 7: Shared expertise, communication, and collaboration around national data

Data advocates are facing a major gap in public perception about the value that government data provides. Most Americans don't think about how data impacts their daily lives, such as the ways it influences access to healthcare, public transportation design, or responses to wildfires in their communities. And if they do, they may assume that the data that has historically existed will still be reliable in the future. Building broad public support for the nation's data ecosystem is essential to ensuring its long-term durability.

Roundtable participants were acutely aware of the need to find new ways to talk about data's value. Many noted that data is vital infrastructure whose role generally goes unseen. Others recognized that the people who are closest to the data may not be the best ones to rally support for it; as one said, "it is really important for nerdy information to come from non-nerds." The overall goal, participants agreed, is to "make data matter to people through creative engagement," which requires a specific set of skills.

> **There are a lot of people who aren't statisticians or data scientists, but who rely on federal datasets to determine when to plant their crops or where to start a new business. When that federal data goes away, it's important to help everyone impacted understand how to organize, how to advocate, how to story tell around how the loss of that data makes their work harder. Those constituencies may have more sway over getting things back in place and have compelling stories to share.**
>
> Lynn Overmann, Beeck Center

Ongoing priorities

Tell compelling stories about the value and credibility of federal data

The concept of data storytelling is central to communicating the value of data. Several Roundtable discussions explored different aspects of effective storytelling, including:

Using intermediaries

It's clear that government data providers may not be the best storytellers about the value of the data they produce — or be adequately funded to do so for all audiences. Effective storytelling intermediaries can include journalists, NGOs, social media influencers, academic and other institutions, businesses, and others. Organizations like USAFacts that act as intermediaries between the public and government data can develop easy-to-understand data summaries and visuals for public awareness and education.

Highlighting local applications

Promoting the value of local applications could help make data more relevant to the public. Data on issues as mundane as road closures and bus schedules, or as significant as municipal budget analysis, can show the range of uses for both state and local data and federal data applied at a local level.

Using data visualizations

As one participant noted, "Value increases when data is made legible, not just made public." Visualization and interpretation are critical tools to communicating the meaning of data. In addition, these kinds of data publication can take care to attribute the sources of data so that users can make the connection between the data they're seeing and the agencies or entities that provided it.

Engage with the public around key government data programs

Federal data ultimately serves the public, and the public is an important voice in setting priorities for its future. Agencies need input and feedback from businesses, nonprofit organizations, researchers and academics, journalists, community-based organizations, state and local governments, and a wide range of other data users.

OMB's guidance on implementing the Evidence Act states that agencies must "consider whether the public has provided feedback about the value of the public data asset. For example, uses that are in the public interest may include news reporting, policy development, and research in the physical sciences, social sciences, humanities, or another field."[80,81] The memo requires each agency to establish "processes and procedures...to facilitate collaboration with stakeholders outside the Federal Government...for the purpose of understanding how data users value and use government data."[82]

Many aspects of federal data policy already require changes to be published in the Federal Register, and agencies can also issue Requests for Information (RFIs) for broad public input. But agencies can follow the OMB directive in ways that go well beyond these standard actions, through town halls, roundtables, workshops, or other forms of outreach. CODE's Open Data Roundtables are one model, and CODE and the Data Foundation have published a Stakeholder Engagement Toolkit for this work.[83] Data advocates can also convene independent public discussions on federal data, like the recently launched Federal Data Forum.[84]

Emerging opportunities

Target communications to key audiences

While broad public communication is valuable, it's not the only way, or necessarily the best way, to rally support for critical data resources. Targeted outreach to key data stakeholders and decision-makers could focus more effectively on people who have a special strong interest in the data or the ability to influence how data is collected, managed, and used. Two kinds of mapping exercises can provide a basis for targeted outreach to advocate for better national data.

Stakeholder mapping

A formal, intensive project could analyze the different groups with a stake in national data and the best ways to work with them. This analysis could include key players in each sector, an analysis of any relevant public positions they have taken, key contacts, relative value of data to their work, key relevant datasets, and more.

Power mapping of decision-makers

This project would use "power mapping," a common tool for advocacy and social justice organizations, to identify individuals and entities with the greatest ability to influence and improve national data and the issues of importance to them. These could range from state and local government leaders to U.S. Senators and Representatives. The mapping could, for example, include analysis of the importance of certain kinds of data, or data-producing research projects, in different Congressional districts, drawing on the kind of work now being done by the Impact Project.

Develop and tap expert networks of former federal workers

The restructuring of the federal workforce in 2025 included numerous changes to agencies that collect federal data. Many Roundtable participants flagged this as a cause for concern. As one put it, "every part of the government that touches data has experienced significant losses of expertise."

There is an opportunity to tap former federal experts to help navigate new reforms to the federal data ecosystem. In addition to utilizing the knowledge of experts who have been recently laid off, this approach could create networks of former federal experts who could advise on new data initiatives for years to come. A possible model is the Environmental Protection Network, which for years has maintained a volunteer network of former employees of the Environmental Protection Agency (EPA).

Another strategy is to tap former feds to create non-governmental data advisory committees. Early in his administration, President Trump reinstated an executive order from his first term, instructing all federal agencies to reduce the number of their official Federal Advisory Committees by at least one-third.[85] In cases where important government advisory committees have been disbanded, new, outside committees can review the health of federal data and advise on data programs from outside of government. As an example, in August 2025, the Census Scientific Advisory Committee (CSAC) announced that it was reconstituting itself outside of government under the name I-CSAC, with the "I" standing for Independent.[86] The Union of Concerned Scientists has published a toolkit to help others create similar committees.[87,88]

CONCLUSION

The path forward — building an ongoing data community

Conclusion: The path forward — building an ongoing data community

> **I love data literacy and democratizing data. You can increase data skills, you can also decrease data barriers, and both of those have to happen at the same time. The federal government isn't the only actor here, though. There's a critical infrastructure of civil society and the private sector that can help in that last mile of turning data into action.**
>
> Denice Ross, former U.S. Chief Data Scientist

As the United States approaches its 250th birthday, an uncomfortable truth is emerging: America doesn't have a national data system — it has a collection of parts. For more than two centuries, federal data has grown agency by agency without a coherent design. That system still powers everything from GDP to public health, but it's now under strain, facing political disruption, declining survey reliability, outdated infrastructure, and new demands from AI.

Through two Roundtables and many discussions, USAFacts and CODE brought together a bipartisan group of experts with diverse backgrounds but similar priorities. Virtually all agreed that federal data should be treated as core national infrastructure, governed through a modern, secure, AI-ready data ecosystem. This broad consensus reflected decades of bipartisan agreement, where legislation to improve America's data has passed Congress by near-unanimous votes.

The Roundtable participants also confirmed that America's national data is at a transitional moment. They saw unprecedented risks to the country's core data, opportunities for building and improving the nation's data resources, and strong needs for a coherent vision for the future, coordinated action, and coalition-building. Better coordination can reduce duplication, strengthen communication networks, and catalyze action. It can help public, private, and civic actors collaborate to build a new national data ecosystem — one that reflects democratic principles, including transparency, inclusion, and accountability.

We believe that the goals described in this paper, which derive from the insights of more than 150 data experts, can provide a starting point of reference for the collaborative approach that is needed today. Those goals are:

- Maintaining critical data collections.
- Strong, reliable federal data governance.
- A national data ecosystem that serves all Americans.
- An AI-enabled system for better data and data access.
- Extensive and usable systems of state and local data.
- Robust American data developed with and for America's businesses.
- Shared expertise, communication, and collaboration around national data.

We view these seven goals as essential to creating a national data ecosystem built for the 21st Century.

We need to recognize the critical importance of America's data infrastructure and acknowledge that it needs to improve precisely because it is essential. The challenge facing us isn't political, but foundational. The question is whether the U.S. will build a data system strong enough to last to its 300th birthday or continue relying on one that was never designed to get us there. We already have broad bipartisan consensus on what the country needs. Now we need the political will, and the sense of public urgency, to build the national data ecosystem America deserves.

Acknowledgements

This paper was planned, developed, and published in collaboration by USAFacts and CODE. CODE's president, Joel Gurin, drafted the paper with editorial support and input from Matt Rumsey and support from Paul Kuhne and Temilola Afolabi at CODE. Richard Coffin and Ben Gitis at USAFacts provided guidance, policy insight, editorial review, and extensive substantive input, together with their colleagues Lauren Woodman, Kari D'Elia, Jessica Piha, Pieter van Noordennen, and Megan Winfield. USAFacts and CODE would like to thank the Advisory Group for their Roundtable program: Meeta Anand, Gary Bass, William Beach, Brittany Janis, Daniella Lowenberg, Amy O'Hara, Steve Pierson, Alonzo Plough, Denice Ross, and Jed Sundwall. We also thank the following individuals who contributed insights, information, and review at different stages of the paper's development: Jennifer Bell-Ellwanger, Daniel Chenok, Christopher Dick, Amy Edwards Holmes, Flo Gutierrez, Stephen Hardy, Cameron Hickey, George Hobor, Alex Howard, Brenna Isman, Beth Jarosz, Barry Johnson, Meghan Maury-Fox, and Alan Shark. These experts advised on this paper as individuals rather than as representatives of their organizations, and their inclusion on this list does not imply their endorsement of this paper's analysis or recommendations.

About USAFacts and CODE

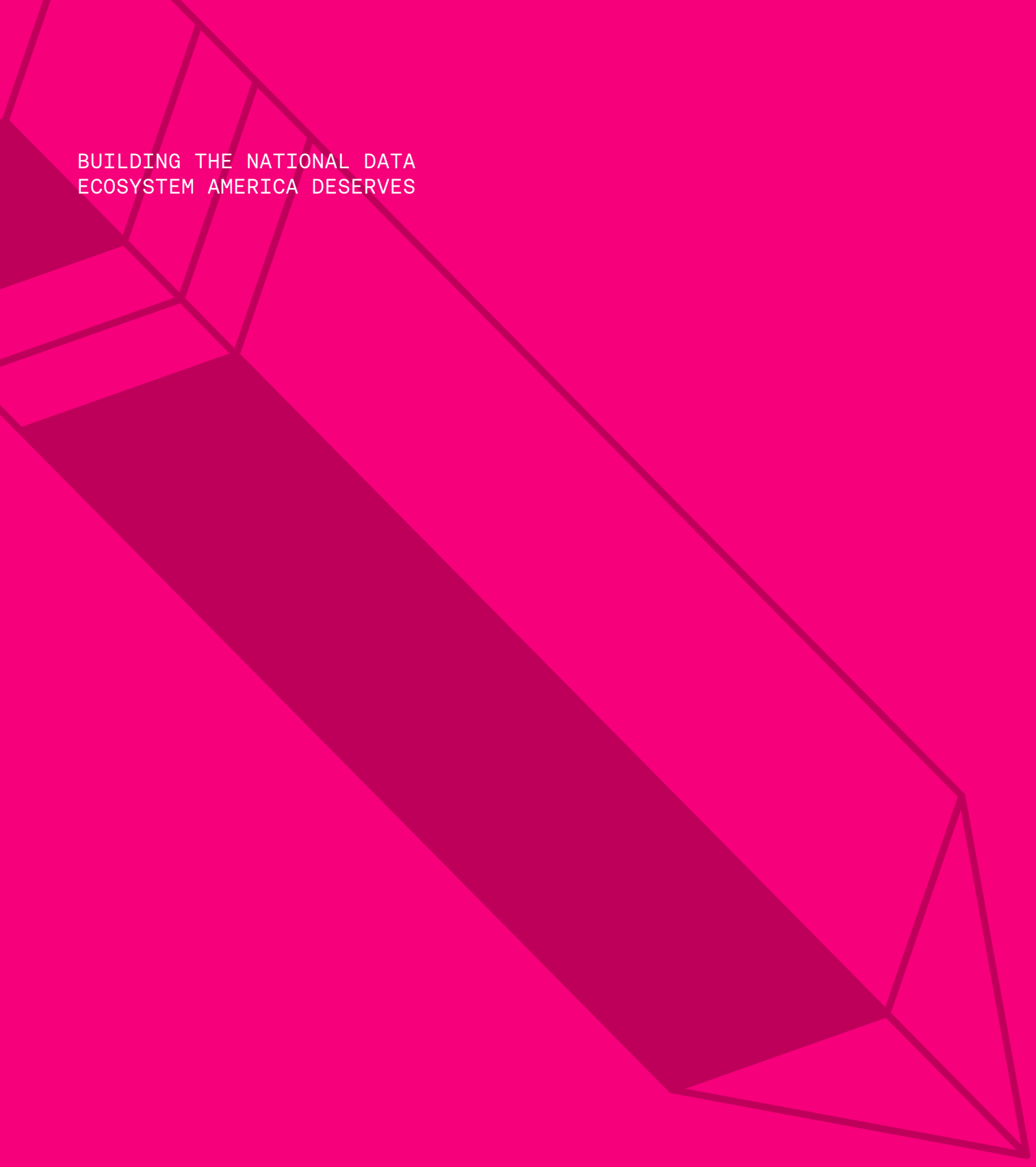
Founded by former Microsoft CEO Steve Ballmer, USAFacts is a not-for-profit, nonpartisan civic initiative that provides a data-driven portrait of the American population, governments' finances, and governments' impact on society. At USAFacts, we believe a vibrant democracy requires informed debate grounded in facts. All Americans are stakeholders in this democracy. To find solutions to issues affecting the United States, everyone, from regular citizens to top policymakers, needs data to understand how the government serves the people. We exclusively use publicly available government data to inform the American public with facts.

We strive for our facts to be unbiased, understandable, comprehensive, contextual, and transparent. We rely only on numbers from government agencies and present them without bias, leaving you to use the data to judge the country's direction for yourself. We gather metrics from government sources and standardize them so they're easy to grasp. We offer a complete view of government finances and impact, from the federal level to the community where you live. We use government data from many different sources, allowing you to see the big picture in one place. Lastly, we are open about our data sources and methods.

The Center for Open Data Enterprise (CODE) is an independent 501(c)3 nonprofit organization based in Washington, D.C. CODE's mission is to harness the power of open and shared data for the public good by working with government agencies, businesses, nonprofits, and researchers who are both data providers and data users. Since it was founded in January 2015, CODE has held numerous Open Data Roundtables with the White House, a dozen different federal agencies, and independent partners in the U.S.; has published a wide range of research papers and online tools; and has implemented several international projects. In addition to working with government agencies and officials, CODE partners with private-sector companies, foundations, and other nonprofit organizations to fulfill its mission. More information is available at CODE's website, www.odenterprise.org.

Creative Commons

BUILDING THE NATIONAL DATA
ECOSYSTEM AMERICA DESERVES



Appendices

Appendix 1: Acronym list

AI	Artificial Intelligence
APA	Administrative Procedure Act
API	Application Programming Interface
APDU	Association of Public Data Users
ASA	American Statistical Association
BEA	Bureau of Economic Analysis
BLS	Bureau of Labor Statistics
CBO	Community-Based Organization
CDC	Centers for Disease Control and Prevention
CDO	Chief Data Officer
CEJST	Climate and Economic Justice Screening Tool
CMS	Centers for Medicare and Medicaid Services
CODE	Center for Open Data Enterprise
D-DAN	Data Disaggregation Action Network
DATA Act	Digital Accountability and Transparency Act
DOGE	Department of Government Efficiency
DOI	Digital Object Identifier
DSA	Data Sharing Agreement
EO	Executive Order
EPA	Environmental Protection Agency
FAIR	Findable, Accessible, Interoperable, and Reusable
GenAI	Generative AI
I-CSAC	Independent Census Scientific Advisory Committee
IPEDS	Integrated Postsecondary Education Data System
LLM	Large Language Model
MCP	Model Context Protocol
MOU	Memorandum of Understanding
NAIRR	National Artificial Intelligence Research Resource
NCSES	National Center for Science and Engineering Statistics
NGO	Non-Governmental Organization
NOAA	National Oceanic and Atmospheric Administration
NSDS	National Secure Data Service
NSF	National Science Foundation
NSO	National Statistics Office
NTD	Nontraditional Data
OECD	Organization for Economic Cooperation and Development
OIRA	Office of Information and Regulatory Affairs
OMB	Office of Management and Budget
PEDP	Public Environmental Data Partners
PRA	Paperwork Reduction Act
RAG	Retrieval-Augmented Generation
RFI	Request for Information
RWJF	Robert Wood Johnson Foundation
SEAP	Southern Economic Advancement Project
SOGI	Sexual Orientation and Gender Identity
SPD 15	Statistical Policy Directive 15
USGS	U.S. Geological Survey

Appendix 2: Roundtable participating organizations (Washington and Oakland combined)

- American Action Forum
- American Enterprise Institute
- American Governance Institute
- American Institutes for Research (AIR)
- American Millennium Society
- American Statistical Association
- Annie E. Casey Foundation
- Arizona State University (ASU)
- Associated General Contractors of America
- Association of Public Data Users
- Athena Civic Consulting
- Basys.ai
- Beeck Center for Social Impact + Innovation
- Bill & Melinda Gates Foundation
- Bipartisan Policy Center
- Bloomberg Center for Government Excellence
- BrightQuery
- Brookings Institution
- California Department of Technology Office of Digital Services
- California Digital Library
- California Health Benefits Review Program
- Cato Institute
- Center for Democracy and Technology
- Center for Open Science
- Center for Regional Economic Competitiveness
- Center on Privacy & Technology @ Georgetown Law
- City Health Dashboard and Congressional District Health Dashboard
- City of San Leandro
- Civic Texts
- Civic Wisdom
- CNSTAT, National Academies of Sciences, Engineering, and Medicine
- Committee on National Statistics
- Common Health Coalition
- Cornell University ILR School
- CSIS
- Data Foundation
- Data Funders Collaborative
- Data Quality Campaign
- Data Rescue Project
- dataindex.us
- Delivery Associates
- Democracy Fund
- Demographics Analytics Advisors
- EarthJustice
- Environmental & Health Data & Analysis Trust (EHDAT)
- Environmental Data & Governance Initiative
- Environmental Law Institute
- Environmental Policy Innovation Center (EPIC)
- Esri
- Federation of American Scientists
- Fiscal Lab on Capitol Hill
- GDB Consulting
- Georgetown University
- Georgetown University - Massive Data Institute
- Goldman School of Public Policy, UC Berkeley
- Google
- GovLab
- Hillcrest Advisory
- IBM Center for the Business of Government
- Informatics for Health
- Institute for Excellence in Government
- International Association for Research in Income and Wealth
- Internet Archive
- Intertidal Agency
- Kapor Foundation
- Maryland Department of Information Technology
- Measures for Justice
- Mission Measurement/Impact Genome
- Murmuration

- National Academies of Science
- National Conference on Citizenship
- National League of Cities
- Niskanen Center
- North Carolina Institute of Climate Studies
- Open Environmental Data Project
- Overdeck Family Foundation
- ParadigmAnalysis
- Partnership for Public Service
- Pew Research Center
- Population Association of America
- Population Reference Bureau
- Portland State University
- Project Evident
- Prolegis
- Public Environmental Data Partners
- Public Good Group
- Public Technology Institute
- Radiant Earth
- Recoding America Fund
- Results for America
- Reveal Global Consulting
- Robert Wood Johnson Foundation
- Science Philanthropy Alliance
- SF Marin Food Bank
- Southern Economic Advancement Project
- SSRS
- Stand Together
- Stanford Institute for Human-Centered AI
- Stanford University Library (SUL)
- State of California
- The Brookings Institution
- The Heritage Foundation
- The Impact Project
- Transparent Technology
- Trust for Public Land
- UC Berkeley
- UC Office of the President
- UC-Berkeley, Schmidt Center for Data Science and Environment
- University of Colorado Health Sciences Center
- Urban Institute
- USAFacts
- Walton Family Foundation
- Washington State Library (OSOS)
- Washington Technology Solutions (WaTech), State Data Office, Open Data Program
- West Sacramento Police Department
- Yale Tobin Center for Economic Policy

Appendix 3: Washington, DC Roundtable agenda

Agenda: Roundtable on Building the National Data Ecosystem America Deserves

October 15, 2025, 9:00 a.m. - 5:00 p.m. EST | Reception: 5:00 - 7:00 p.m. EST

600 14th St NW, Suite 600, Washington, DC 20005, United States

8:45 am: Registration and Light Breakfast

9:30 am: Welcome Remarks

Richard Coffin, Chief of Advocacy and Research, USAFacts

9:40 am: Plan for the Day

Joel Gurin, President, Center for Open Data Enterprise (CODE)

9:50 am: Panel Discussion

The Value and Future of Federal Data

- Moderator: Ben Gitis, Data Policy and Advocacy Manager, USAFacts
- Denice Ross, Senior Fellow, Federation of American Scientists
- Erica Groshen, Senior Economics Advisor, Cornell University School of Industrial and Labor Relations
- Fred Ashton, Director of Competition Policy, American Action Forum

10:35 am: Breakout Session 1

Engaging Around Federal Data

11:45 am: Break

12:00 pm: Panel Discussion

Data Modernization, Governance, and Trust

- Moderator: Joel Gurin
- Victoria Houed, Managing Director, Recoding America Fund
- Jennifer Huddleston, Senior Fellow in Technology Policy, Cato Institute
- Stefaan Verhulst, Co-founder of the Governance Laboratory (The GovLab)

12:40 pm: Lunch

1:30 pm: Breakout Session 2

Data Modernization, Governance, and Trust

2:20 pm: Break

2:30 pm: Panel Discussion

Filling Federal Data Gaps: State, local, and other major data sources

- Moderator: Daniella Lowenberg, Director, Data Funders Collaborative
- Lynn Overmann, Executive Director, Beeck Center for Social Impact + Innovation, Georgetown University
- Chris Jackson, Senior Vice President - Civil Society Division Lead, SSRS
- Afia Tyus, Research Associate, Southern Economic Advancement Project (SEAP)

3:00 pm: Breakout Session 3

Filling Federal Data Gaps

4:00 pm: Break

4:15 pm: Reports back from the full day

4:50 pm: Closing Remarks

Joel Gurin and Richard Coffin

5:00 pm: Reception at the Roundtable venue

Appendix 4: Oakland, CA Roundtable agenda

Agenda: Roundtable on Building the National Data Ecosystem America Deserves

November 6, 2025, 9:00 a.m. - 5:00 p.m. PST | Reception: 5:00 - 7:00 p.m. PST

The Kissel Uptown Oakland, 2455 Broadway, Oakland, CA 94612

9:00 am: Registration and Light Continental Breakfast

9:30 am: Welcome Remarks

Richard Coffin, Chief of Advocacy and Research, USAFacts

9:40 am: Plan for the Day

Joel Gurin, President, Center for Open Data Enterprise (CODE)

9:45 am: Opening Speaker

Bill Beach, Former Commissioner of Labor Statistics

10:00 am: Panel Discussion

A Future National Data Ecosystem: Goals, Stakeholders, and Governance Needs

- Moderator: Moderator: Ben Gitis, Data Policy and Advocacy Manager, USAFacts
- Beth Jarosz, Vice President, Association of Public Data Users (APDU)
- Steve Diggs, Research Data Specialist and Sr. Product Manager, UC Office of the President
- Nick Hart, President and CEO, The Data Foundation

10:45 am: Breakout Session 1

Engaging to Shape the Future of National Data

11:45 am: Break

12:00 pm: Panel Discussion

The Need for State and Local Data

- Moderator: Joel Gurin
- Mayor Juan Gonzalez, City of San Leandro
- Jonathan Porat, Chief Technology Officer, State of California
- Cathi Greenwood, Open Data Program Manager, Washington Technology Solutions
- Rob Goldman, CEO, Transparent Technology

12:30 pm: Lunch

1:30 pm: Breakout Session 2

State and Local Government in a new National Data Ecosystem

2:20 pm: Break

2:30 pm: Panel Discussion

Modernizing National Data with Emerging Technology

- Moderator: Seth Stephens-Davidowitz, Senior Advisor, USAFacts
- Jed Sundwall, Executive Director, Radiant Earth
- Kari D'Elia, Chief Product Officer, USAFacts
- Simon Rogers, Data Editor, Google

3:00 pm: Breakout Session 3

Emerging Technology for a Stronger National Data Ecosystem

4:00 pm: Break

4:15 pm: Reports back from the full day

4:50 pm: Closing Remarks

Joel Gurin and Richard Coffin

5:00 pm: Reception at the Roundtable venue

Appendix 5: A timeline of data legislation and policy

1934 National Archives created to centralize federal record-keeping.[89]

1946 Administrative Procedure Act (APA) enacted; includes processes for data collections.[90]

1974 The Privacy Act establishes fair information practices for the collection, management, and use of individuals' information from federal records.[91]

1980 Paperwork Reduction Act (PRA) lays out guidelines for federal information collections.[92]

1995 PRA is revised with added requirements for agencies to seek public comment on proposed information collections and then have them reviewed by the White House OMB.[93]

2000 Information Quality Act is enacted to ensure the quality, objectivity, utility, and integrity of statistical and other information from federal agencies.[94]

2009 President Barack Obama issues a memorandum on transparency and open government on day one of his presidency.[95]

2009 Launch of data.gov, the federal government's centralized portal to all open government data.[96]

2013 President Obama issues an executive order (EO) to make open and machine-readable data the new default for government information.[97] OMB simultaneously releases an Open Data Policy that builds on the administration's open data initiatives already underway.[98,99]

2014 The Digital Accountability and Transparency Act of 2014 (the DATA Act) is enacted to establish new standards for the transparency of federal spending data, made available through the website USASpending.gov.[100,101]

2018 President Trump signs the Foundations for Evidence-Based Policymaking Act (the Evidence Act), which sets requirements for all federal data.[102]

2020 President Trump continues the government's commitment to open data by launching the Federal Data Strategy.[103] The Strategy, developed through a collaborative effort involving 23 agencies, is designed to guide annual action plans over the following decade to "leverage data as a strategic asset."[104]

2025 In the last week of the Biden administration, OMB issues a long-awaited memorandum with specific guidance on open government data access and management under the Evidence Act.[105]

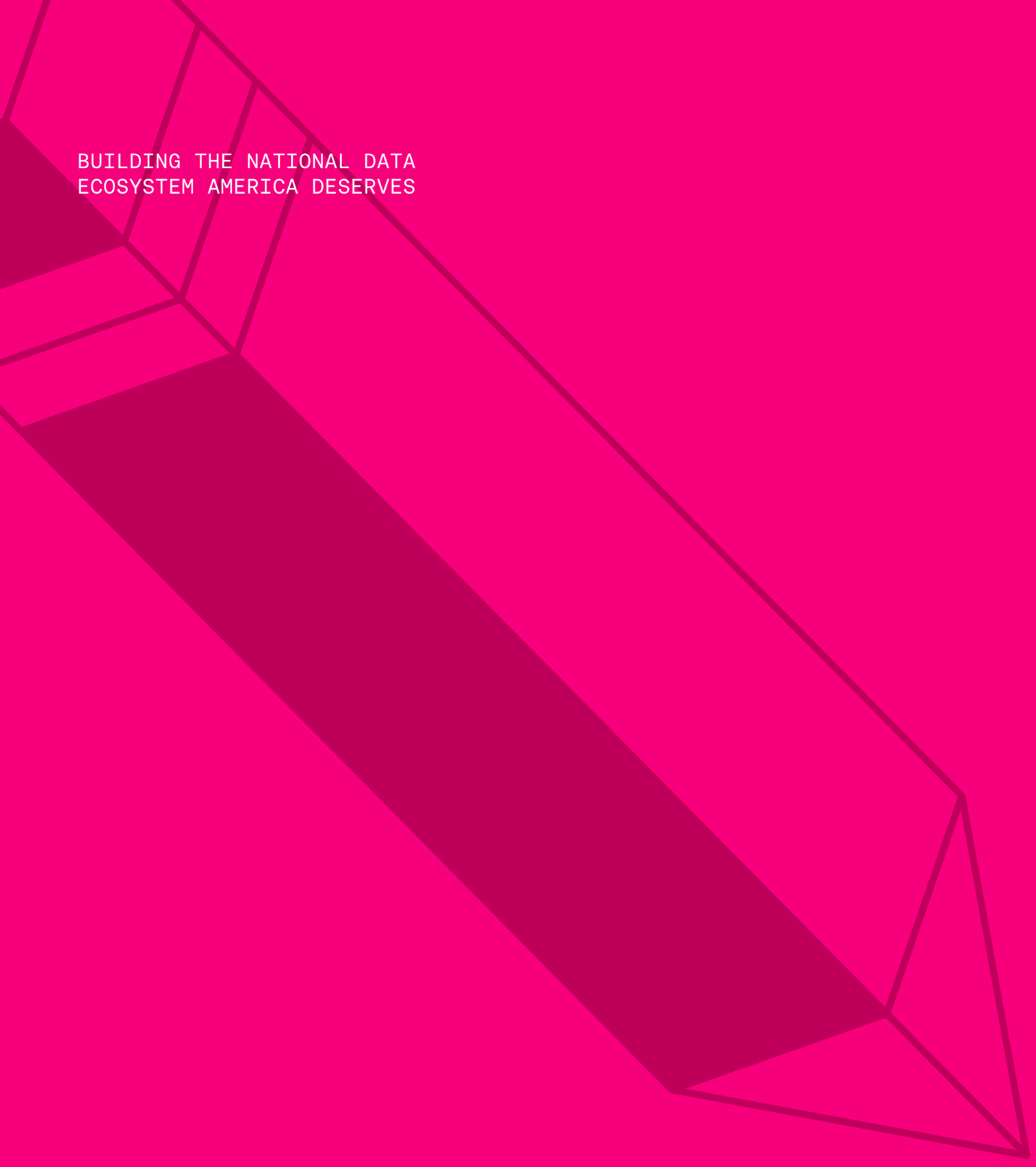
Appendix 6: Discontinued federal datasets

There is no definitive, continually updated list of all the federal datasets that have been discontinued since January 2025. The best source is the list of "Confirmed Data Terminations and Removals"[106] published by America's Essential Data in collaboration with the Federation of American Scientists, Public Environmental Data Partners, and The Impact Project. The published list, which was last updated in December 2025, lists the following ten terminated datasets:

- Agricultural Labor Survey (USDA)
- Community Resilience Estimates Equity Supplement (Census)
- Current Population Survey Food Security Supplement (USDA)
- Demographic and Health Surveys (USAID)
- Development Data Library (USAID)
- Drug Abuse Warning Network (SAMHSA)
- Federal Employee Viewpoint Survey (Office of Personnel Management)
- Greenhouse Gas Reporting Program (EPA)
- National Beneficiary Survey (Social Security Administration)
- NOAA Research and Development Database (NOAA)

In addition, the website notes the government-wide removal of sexual orientation and gender identity data, lists several datasets that have had other variables removed, and lists tools for accessing data that have been discontinued, including the Climate and Economic Justice Screening Tool (CEJST) and the EPA's EJScreen, and NOAA's Billion-Dollar Weather and Climate Disasters database.

A dataset does not have to be completely eliminated to become essentially unavailable to the public. For example, the Department of Homeland Security's Homeland Infrastructure Foundation-Level Data (HIFLD) datasets have been moved from an open, publicly accessible portal to restricted access. In other cases, like the Pregnancy Risk Assessment Monitoring System (PRAMS), the federal staff responsible for aggregating and analyzing state data has been decimated or eliminated. Even if state-level data is still collected, it will now be impossible to use it to develop a national database.



Endnotes

1. "Epilogue: Securing the Republic: James Madison to W. T. Barry." University of Chicago Press. https://press-pubs.uchicago.edu/founders/documents/v1ch18s35.html.
2. Bureau, US Census. 2022. "1790 Census." U.S. Census Bureau. January 18, 2022; data.gov. https://www.census.gov/library/photos/1790_census.html.
3. https://usafacts.org/.
4. https://www.opendataenterprise.org/.
5. https://www.opendataenterprise.org/what-we-do#methodology.
6. Available at https://bit.ly/NationalDataResources.
7. "APDU Guiding Principles for Public Data – Association of Public Data Users." APDU.org. Updated March 2026. https://apdu.org/apdu-guiding-principles-for-public-data/.
8. Remarks at the November 2025 USAFacts/CODE Roundtable in Oakland, CA. Quotes from Roundtable speakers and panelists are taken verbatim from recordings or edited with the speaker's permission.
9. "What is Open Data?" *Open Data Handbook.* https://opendatahandbook.org/guide/en/what-is-open-data/.
10. *America's Essential Data.* https://essentialdata.us/.
11. https://www.uschamberfoundation.org/workforce/public-private-standards-development-and-use-by-government.
12. Anonymous quotes from participants are paraphrased from notetakers' reports of the breakout sessions.
13. https://essentialdata.us/, https://datafoundation.org/, https://datafoundation.org/pages/the-people-s-data-100.
14. https://dataindex.us/.
15. https://www.datarescueproject.org/.
16. https://screening-tools.com/.
17. https://archive.org/.
18. https://www.federaldatafieldguide.us/.
19. https://data.gov/.
20. https://www.opendataenterprise.org/publications#Future-Health-Data.
21. This example was shared at the October 2025 Washington DC Roundtable by former U.S. Chief Data Scientist Denice Ross.
22. "The OPEN Government Data Act." Congressional Research Service. December 29, 2022. https://www.congress.gov/crs_external_products/IF/PDF/IF12299/IF12299.4.pdf.
23. Tanner, Brooke, Josie Stewart, and Stephanie K Pell. "Privacy under Siege: DOGE's One Big, Beautiful Database." Brookings. June 25, 2025. https://www.brookings.edu/articles/privacy-under-siege-doges-one-big-beautiful-database/.
24. "OMB Memos", Evidence Act Hub, The Data Foundation, https://datafoundation.org/news/evidence-act-hub/756/756-OMB-Memos. (accessed February 6, 2026).
25. https://datafoundation.org/pages/evidence-act-hub.
26. https://www.pewresearch.org/internet/2023/10/18/how-americans-view-data-privacy/.
27. In this paper we will generally refer to DSAs, but in some cases MOUs may be a more appropriate approach.
28. "Federal Data Excellence," USAFacts. https://usafacts.org/research-and-initiatives/fde/.
29. https://catalog.data.gov/.
30. "Timmons Leads Bipartisan Push to Reestablish Commission on Evidence-Based Policymaking." *U.S. Representative William Timmons.* May 13, 2025. https://timmons.house.gov/news/documentsingle.aspx?DocumentID=195.

31. "Supreme Court Strikes Down Chevron, Curtailing Power of Federal Agencies," SCOTUSBlog, June 28, 2024, https://www.scotusblog.com/2024/06/supreme-court-strikes-down-chevron-curtailing-power-of-federal-agencies/. (accessed February 6, 2026).

32. Handbook of Statistical Organization, Third Edition: The Operation and Organization of a Statistical Agency, United Nations Statistics Division, 12, accessed August 22, 2025, https://unstats.un.org/unsd/dnss/docViewer.aspx?docID=1519.

33. McLaughlin, Michael. "Consolidating Statistical Agencies Would Lead to Better Data, More Efficient Spending." *Center for Data Innovation.* September 16, 2018. https://datainnovation.org/2018/09/consolidating-statistical-agencies-would-lead-to-better-data-more-efficient-spending/.

34. Office of Management and Budget. "Technical Supplement to the 2026 Budget." *GovInfo,* 2025 https://www.linkedin.com/posts/nickrhart_the-detailed-version-of-the-presidents-fy-activity-7334331417109159938-0ADc/?rcm=ACoAAABMmfkBeLRcYh5pkOwQrbbEah3ut7YXtn4.

35. https://www.whitehouse.gov/wp-content/uploads/2026/04/budget_fy2027.pdf.

36. Wasserstein, Ron. "The Case for Careful Reform of the Federal Statistical System: Preserving Data Integrity in an Era of Government Transformation." *Federal News Network.* May 28, 2025. https://federalnewsnetwork.com/commentary/2025/05/the-case-for-careful-reform-of-the-federal-statistical-system-preserving-data-integrity-in-an-era-of-government-transformation/.

37. "Modernizing the Federal Statistical System." *American Statistical Association, 2025.* https://www.amstat.org/policy-and-advocacy/modernizing-the-federal-statistical-system/.

38. "What does the Bureau of Labor Statistics do?," USAFacts, 2024. https://usafacts.org/explainers/what-does-the-us-government-do/subagency/bureau-of-labor-statistics/.

39. https://thecensusproject.org/, https://www.friendsofbls.org/.

40. Matt Grossman and Harriet Torry, "Fed Chair Jerome Powell Says U.S. May Be Drastically Overstating Job Numbers," *The Wall Street Journal,* December 12, 2025.

41. "The National Secure Data Service Demonstration," NSF/NCSES, https://ncses.nsf.gov/initiatives/national-secure-data-service-demo.

42. Measuring Race and Ethnicity across the Decades: 1790—2010 - U.S. Census Bureau." n.d. www.census.gov. https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html.

43. Alms, Natalie. "LGBTQ Data Is Disappearing under Trump, Reports Find." Nextgov/FCW, March 3, 2026. https://www.nextgov.com/digital-government/2026/03/lgbtq-data-disappearing-under-trump-reports-find/411831/?oref=ng-author-river.

44. https://civilrights.org/edfund/, https://civilrights.org/edfund/data-disaggregation-action-network/.

45. "Disaggregation Nation: State Pathways and Key Considerations for Improved Race and Ethnicity Data," July 2025. https://civilrights.org/edfund/wp-content/uploads/sites/2/2025/07/DisaggregationNation2025.pdf.

46. See, for example, https://www.linkedin.com/company/the-peoples-data-project/.

47. "Generative Artificial Intelligence," *Wikipedia,* https://en.wikipedia.org/wiki/Generative_artificial_intelligence.

48. https://www.whitehouse.gov/wp-content/uploads/2026/03/03.20.26-National-Policy-Framework-for-Artificial-Intelligence-Legislative-Recommendations.pdf.

49. Oliver Wise, Sallie Ann Keller, and Victoria Houed, "Preparing Open Data for the Age of AI," U.S. Department of Commerce, January 18, 2024. https://www.commerce.gov/news/blog/2024/01/preparing-open-data-age-ai.

50. "The USAFacts Guide to AI-Ready Data for Government Agencies." USAFacts. https://media.usafacts.org/m/634ac133d72ded81/original/USAFacts_AIReadinessForGovernment.pdf.

51. "Democratizing Government Data with GenAI: Issues and Opportunities." The Center for Open Data Enterprise. February 2025. https://drive.google.com/file/d/1Nx1YrRdqOGiMvaCMCQZGtvH7xx_0OM7Q/view?pli=1.

52. Verhulst, Stefaan. Zable, Adam. "Unlocking Public Value with Non-Traditional Data: Recent Use Cases and Emerging Trends." *Open Data Policy Lab.* April 9. 2025. https://opendatapolicylab.org/articles/new-blog-post-unlocking-public-value-with-non-traditional-data-recent-use-cases-and-emerging-trends/.

53. Ron S. Jarmin, "Source Data Innovation at the Census Bureau: Improving the Quality and Sustainability of Our Statistics." *U.S. Census Bureau,* August 19, 2025. https://www.census.gov/newsroom/blogs/director/2025/08/source-data-innovation.html.

54. Fabrice Murtin and Max Salomon-Ermel, "Nowcasting subjective well-being with Google Trends: A meta-learning approach," *OECD,* June 28, 2024. https://www.oecd.org/en/publications/nowcasting-subjective-well-being-with-google-trends_cbdfb5d9-en.html; David Kohns and Arnab Bhattacharjee, "Nowcasting growth using Google Trends data: A Bayesian Structural Time Series model," International Journal of Forecasting, July-September 2023. https://www.sciencedirect.com/science/article/pii/S0169207022000620.

55. Emmanuel Asiedu Brempong and Shreya Agrawal, "Nowcasting on Search is bringing AI-powered weather forecasts to people across Africa," Google Africa Blog, March 28, 2025. https://blog.google/intl/en-africa/products/explore-get-answers/nowcasting-on-search-is-bringing-ai-powered-weather-forecasts-to-users-across-africa/.

56. NORC, "Data Concierge Model Report: Two Proposed Data Concierge Models," *America's DataHub Consortium*, November 2024. https://www.americasdatahub.org/wp-content/uploads/2025/03/DCS_23_N03-Concierge-Model-Final-Report_Public.pdf.

57. "ONC - 20 Years." Assistant Secretary for Technology Policy. https://www.healthit.gov/20years/.

58. Interview and email communication with Jennifer Bell-Ellwanger

59. Remarks by Bill Beach, USAFacts/CODE Roundtable, Oakland, CA, November 2025.

60. https://www.bls.gov/wrp/.

61. Lauren J. Young, "A 'shadow CDC' is scrambling to fill gaps in public health data," *Scientific American,* February 3, 2026. https://www.scientificamerican.com/article/states-and-medical-societies-are-stepping-up-to-fill-the-cdcs-data-void/.

62. "Northeast Public Health Collaborative." Wikipedia. Accessed April 29, 2026.

63. https://doh.wa.gov/about-us/west-coast-health-alliance.

64. The Center for Systems Science and Engineering. "COVID-19 Dashboard." Johns Hopkins University of Medicine Coronavirus Resource Center. March 10, 2023. https://coronavirus.jhu.edu/map.html.

65. "US COVID-19 cases and deaths by state." USA Facts. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/.

66. "The U.S. Climate Vulnerability Index," Environmental Defense Fund. https://climatevulnerabilityindex.org/.

67. "Administrative Data." Wikipedia. April 19, 2022. https://en.wikipedia.org/wiki/Administrative_data.

68. "Counting Every Voice: Understanding Hard-To-Count and Historically Undercounted Populations." Census.gov. November 7, 2023. https://www.census.gov/newsroom/blogs/random-samplings/2023/10/understanding-undercounted-populations.html.

69. George, Robert M. "Expert View: Harnessing the Power of Administrative Data in State Government." NORC. April 2025. https://www.norc.org/research/library/harnessing-power-administrative-data-state-government.html.

70. Gurin, Joel. "How Open Data Fuels Key Business Sectors." FedScoop. August 16, 2017. https://fedscoop.com/open-data-fuels-key-business-sectors/.

71. Gurin, Joel. "Leveraging Government Data: How It's Making a Difference." FedScoop. June 28, 2018. https://fedscoop.com/leveraging-government-data-making-difference/.

72. Gilmour, Jonathan. "The Trump Administration Is Sabotaging Your Scientific Data." *The Guardian.* April 21, 2025. https://www.theguardian.com/commentisfree/2025/apr/21/trump-administration-scientific-data-climate.

73. "American Community Survey.' U.S. Census Bureau. http://census.gov/programs-surveys/acs/.

74. "America's Essential Economic and Social Data at Risk -2025 Update the Urgent National Need to Protect and Enhance the American Community Survey Data." The Census Project. 2025. https://thecensusproject.org/wp-content/uploads/2025/03/Census-ACS-Report-2025-v5.pdf.

75. Bloomberg, Michael R. 2026. “In Washington’s War on Data, the Economy and Public Will Lose.” Bloomberg.com. Bloomberg. March 25, 2026. https://www.bloomberg.com/opinion/articles/2026-03-25/michael-bloomberg-trump-assault-on-federal-data-solves-nothing.

76. https://en.wikipedia.org/wiki/Waffle_House_Index; https://www.gasbuddy.com/go/fuel-availability-tracker.

77. https://www.pophive.org/.

78. https://overturemaps.org/.

79. Jed Sundwall, Executive Director of the nonprofit Radiant Earth, has described this model most thoroughly.

80. Young, Shalanda. “Memorandum: Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open Government Data Access and Management Guidance,” Executive Office of the President Office of Management and Budget, January 15, 2025. https://bidenwhitehouse.archives.gov/wp-content/uploads/2025/01/M-25-05-Phase-2-Implementation-of-the-Foundations-for-Evidence-Based-Policymaking-Act-of-2018-Open-Government-Data-Access-and-Management-Guidance.pdf.

81. OMB M-25-05 p. 21

82. OMB M-25-05 p. 27

83. http://reports.opendataenterprise.org/Stakeholder-Engagement-Toolkit.pdf.

84. “Home - Federal Data Forum.” Population Reference Bureau. 2024. https://federaldataforum.prb.org/home.

85. Trump, Donald J, “EO 14148: Initial Recission of Harmful Executive Orders and Actions,” January 20, 2025, https://www.federalregister.gov/documents/2025/01/28/2025-01901/initial-rescissions-of-harmful-executive-orders-and-actions. (accessed February 6, 2026)

86. Allison Plyer. https://www.linkedin.com/feed/update/urn:li:activity:7364724937946820609/. (Accessed August 25, 2025)

87. *Union of Concerned Scientists.* https://www.ucs.org/.

88. Finucane, Melissa., Ekwurzel, Brenda., Frey, H. Christopher. “Running an Independent Science Advisory Committee” *Union of Concerned Scientists.* May 2025. https://www.ucs.org/sites/default/files/2025-05/Scientist%20Advisory%20Committees_final_rev_1.pdf.

89. Wikipedia Contributors. 2019. “National Archives and Records Administration.” Wikipedia. March 30, 2019. https://en.wikipedia.org/wiki/National_Archives_and_Records_Administration.

90. “Administrative Procedure Act.” Wikipedia. September 23, 2023. https://en.wikipedia.org/wiki/Administrative_Procedure_Act.

91. U.S. Department of Justice. “Privacy Act of 1974.” Office of Privacy and Civil Liberties Department of Justice. October 4, 2022. https://www.justice.gov/opcl/privacy-act-1974.

92. “Paperwork Reduction Act (PRA).” n.d. Bureau of Justice Assistance. https://bja.ojp.gov/program/it/privacy-civil-liberties/authorities/statutes/1289.

93. U.S. Congress. Paperwork Reduction Act of 1995. Public Law 104–13. 104th Cong., May 22, 1995. In United States Statutes at Large, vol. 109, 163–185.

94. “Agency Information Quality Guidelines.” The White House. 2024. https://obamawhitehouse.archives.gov/omb/inforeg_agency_info_quality_links.

95. The White House. “Transparency and Open Government” Memorandum For The Heads Of Executive Departments And Agencies. The White House. January 21, 2009. https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.

96. Data.gov. https://data.gov/about/.

97. “Executive Order -- Making Open and Machine Readable the New Default for Government Information.” 2013. Whitehouse.gov. May 9, 2013. https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-.

98. Office of Management and Budget. “Open Data Policy—Managing Information as an Asset.” Memorandum M-13-13. Washington, DC: Executive Office of the President, May 9, 2013. https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

99. “Open Government Initiative.” n.d. The White House. https://obamawhitehouse.archives.gov/open.

100. U.S. Congress, “Digital Accountability and Transparency Act of 2014,” Public Law 113–101, 113th Cong., May 9, 2014, in United States Statutes at Large 128 (2014): 1146–1163, https://www.congress.gov/113/bills/s994/BILLS-113s994enr.pdf.

101. “About the Data Transparency Program | TFX: Treasury Financial Experience.” 2016. Treasury.gov. 2016. https://tfx.treasury.gov/data-transparency/about.

102. Ryan, Paul D. 2019. “H.R.4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act of 2018.” U.S. Congress. January 14, 2019. https://www.congress.gov/bill/115th-congress/house-bill/4174.

103. “2020 Action Plan - Federal Data Strategy.” n.d. Strategy.data.gov. https://strategy.data.gov/action-plan/.

104. “Background - Federal Data Strategy.” n.d. Strategy.data.gov. https://strategy.data.gov/background/.

105. Office of Management and Budget, Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open Government Data Access and Management Guidance, Memorandum M-25-05 (Washington, DC: Executive Office of the President, January 15, 2025), https://bidenwhitehouse.archives.gov/wp-content/uploads/2025/01/M-25-05-Phase-2-Implementation-of-the-Foundations-for-Evidence-Based-Policymaking-Act-of-2018-Open-Government-Data-Access-and-Management-Guidance.pdf.

106. "In Memoriam: America's Essential Data." Essentialdata.us, December 1, 2025. https://essentialdata.us/in-memoriam.html.