United States Government Accountability Office

Report to Congressional Committees

September 2025

HIGHLIGHTS OF A FORUM

Expert Views on the Federal Statistical System

GAO-25-107124

GAO Highlights

HIGHLIGHTS OF A FORUM

Expert Views on the Federal Statistical System

GAO-25-107124 September 2025

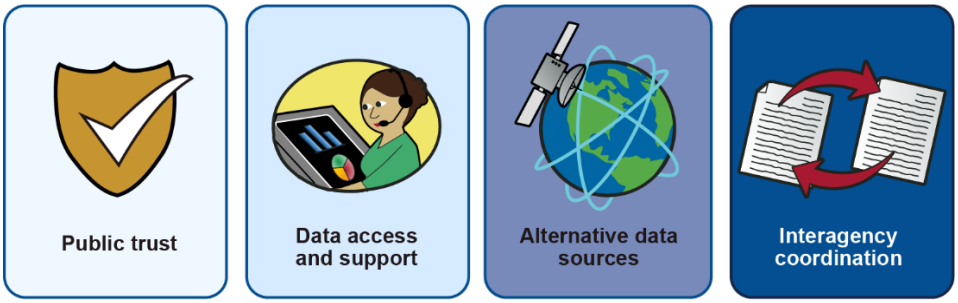
A report to congressional committees.

For more information, contact: Jared Smith at SmithJB@gao.gov

What GAO Found

The federal statistical system faces a critical juncture as it works to modernize and adapt to a rapidly changing data landscape, driven by increasing demand for timely, detailed, and relevant information amid declining survey response rates and rising data collection costs. During a forum GAO held in 2024, experts and stakeholders identified various challenges and opportunities facing the system across a range of topics (see figure).

Figure: Key Topics Discussed at the Forum on the U.S. Statistical System



Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Public Trust. The system faces growing challenges in building and maintaining public trust, particularly as it navigates emerging risks to privacy and confidentiality, according to participants. Suggestions for improving public trust include promoting transparency and advancing privacy enhancing technologies.

Data Access and Support. According to participants, the system faces challenges in meeting the needs of a diverse user base, from highly technical researchers to non-technical data users, and in facilitating access to data products. The system also faces challenges in offering appropriate guidance and tools tailored to users. Potential options for addressing these challenges include expanding data user outreach and training, as well as developing a streamlined data access portal with enhanced analytic capabilities and support.

Alternative Data Sources. Participants highlighted key benefits that alternative data sources—such as private sector data and administrative records—offer for improving federal statistical production and better meeting the needs of data users. Yet participants said that statistical agencies face significant challenges in using alternative data, including legal barriers and dependance on data providers. Participants said that addressing these issues will require strong data security and incentives for provider participation, among other things.

Interagency Coordination. Participants identified effective interagency coordination as key for modernizing statistical production, facilitating outreach to users, and alleviating resource constraints. However, the decentralized design of the system and the absence of a shared framework for interagency data sharing hinder coordination among agencies, creating barriers to data sharing. Suggestions for strengthening interagency coordination include modernizing legislation and establishing shared data infrastructure.

Why GAO Did This Study

The federal statistical system includes 16 statistical agencies and units and over 100 statistical programs that produce data critical for program design, monitoring, and evaluation of federal programs. These data are vital for decisions that directly affect the public. These include the allocation of federal funding to states and localities and the production of key national statistics on health, demographics, and the economy. However, the system faces long-standing challenges that may prevent these agencies from effectively producing timely and accurate information.

In August 2024, GAO held a forum on the federal statistical system. The participants discussed what factors affect the system's ability to (1) build and maintain public trust, (2) meet the needs of its users, (3) sustain and modernize its data collection, and (4) engage in effective interagency coordination. This report is the first in a body of work to assess opportunities to reduce fragmentation, overlap, and duplication in the system, consistent with a statutory provision for GAO to, among other things, routinely investigate government programs to identify duplicative goals and activities.

Participants included 29 experts and stakeholders from the federal statistical system, other federal agencies, state and local government agencies, a non-U.S. national statistical office, an international organization, academic institutions, the private sector, and professional organizations. GAO also interviewed officials from federal and state government agencies. Participants reviewed a draft of this report, and comments were incorporated as appropriate. Views expressed during the proceedings do not represent the opinions of all participants, their affiliated organizations, or GAO.

Contents

Figures

Abbreviations

CIPSEA	Confidential Information Protection and Statistical Efficiency Act
NSDS	National Secure Data Service
OMB	Office of Management and Budget

U.S. GOVERNMENT ACCOUNTABILITY OFFICE
441 G St. N.W.
Washington, DC 20548

September 24, 2025

Congressional Committees

The federal statistical system includes 16 statistical agencies and units, and approximately 100 statistical programs, tasked with providing data on a range of issues, including agricultural production, education, employment, disaster planning, immigration, and other federal policy areas. The system supports decision-making and policy-setting inside and outside the government by providing accurate, objective data. For example, businesses and individuals rely on statistical products to guide financial and life decisions, from making business investments to choosing where to live or work. Key federal statistical products on inflation, unemployment, interest rates, and mortgage rates help Americans and policymakers, such as the Federal Reserve Open Market Committee, understand the economy and plan accordingly.

The decentralized structure of the system—with over a dozen statistical agencies and units operating under the coordination of the Office of Management and Budget (OMB)—creates both opportunities and challenges. Decentralization may encourage individual statistical agencies to tailor data collection and dissemination to the needs of their subject-matter domains and user communities. Because statistical agencies operate under different departments, they are positioned to be responsive to specific policy, operational, and research priorities within their sectors.

However, decentralization and fragmentation across agencies increases the risk of overlapping efforts, duplication in data collection or dissemination activities, and inefficient resource use.[1] Our prior work has addressed issues relevant to the system, highlighting challenges such as fragmented data collection, barriers to data sharing, lack of coordination

[1]GAO, Fragmentation, *Overlap, and Duplication: An Evaluation and Management Guide*, GAO-15-49SP (Washington, D.C.: April 2015).

and capacity constraints.[2] In February 2012, we stressed the importance of interagency collaboration and data governance within the system to produce timely and high-quality statistical information.[3]

Reforms such as the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act) have encouraged the system to reevaluate traditional approaches to data collection, product development, and service delivery in response to declining public participation in surveys, rising demand for granular and policy-relevant data, and growing needs for aggregated data and microdata[4] that serve a broader range of analytic uses.[5] The Evidence Act improves the federal statistical system by creating a framework for federal agencies to take a more comprehensive

[2]GAO, *U.S. Territories: Coordinated Federal Approach Needed to Better Address Data Gaps*, GAO-24-106574 (Washington, D.C.: May 2024); *Work Arrangements: Improved Collaboration Could Enhance Labor Force Data*, GAO-24-105651 (Washington, D.C.: December 2023); *Science and Engineering Statistics: Improved Communication Needed with Stakeholders on Data Needs*, GAO-23-106361 (Washington, D.C.: September 2023); and *Tax Equity: Lack of Data Limits Ability to Analyze Effects of Tax Policy on Households by Demographic Characteristics*, GAO-22-104553 (Washington, D.C.: May 2022).

[3]GAO, *Federal Statistical System: Agencies Can Make Greater Use of Existing Data, but Continued Progress Needed on Access and Quality Issues*, GAO-12-54 (Washington, D.C.: February 2012). We recommended that OMB update guidance on survey and statistical information collection to identify duplication and improve external consultation; enhance dissemination of interagency statistical committee outputs; and issue comprehensive guidance on administrative data sharing. While OMB has addressed the dissemination and data sharing recommendations through the publication of *A Framework for Data Quality* in 2020 and Memorandum M-14-06 in 2014, respectively, it has not yet updated guidance on survey and statistical information collection to reduce duplication. See Federal Committee on Statistical Methodology, *A Framework for Data Quality* (Washington, D.C, Office of Management and Budget, September 2020) and Office of Management and Budget, Guidance for Providing and Using Administrative Data for Statistical Purposes, M-14-06 (Washington, D.C.: Feb. 14, 2014).

[4]Aggregated datasets provide summary information at various levels of grouping. For example, aggregated data could summarize insurance coverage by state, which protects the confidentiality of individual responses. Microdata refers to individual-level records containing detailed information. Unlike aggregate statistics, which summarize data across groups, microdata provides the raw, anonymized responses of individuals or units (such as households, persons, or businesses).

[5]Pub. L. No. 115-435, 132 Stat. 5529 (2019). In 2023, we reported that the effective implementation of the Evidence Act requires strengthened interagency collaboration, transparency, and capacity-building to support the generation and use of evidence across federal agencies. See GAO, *Evidence-Based Policymaking: Practices to Help Manage and Assess the Results of Federal Efforts,* GAO-23-105460 (Washington, D.C.: July 2023).

and integrated approach to evidence building, and enhancing the federal government's capacity to undertake those activities.

In October 2024, OMB issued the *Fundamental Responsibilities of Recognized Statistical Agencies and Units* rule, often referred to as the Trust Regulation, as part of its responsibilities under the Evidence Act, effective December 10, 2024.[6] The Trust Regulation aims to enhance public trust by codifying the roles, responsibilities, and autonomy of Recognized Statistical Agencies and Units ("statistical agencies"), among other things.[7] These new efforts have the potential to address challenges facing the system by expanding data sharing, access, and collection for statistical purposes.[8]

The Statutory Pay-As-You-Go Act of 2010 included a provision for GAO to conduct routine investigations to identify programs, agencies, offices, and initiatives with duplicative goals or activities governmentwide.[9] Consistent with these provisions, GAO has initiated a body of work to assess opportunities to reduce fragmentation, overlap, and duplication in the federal statistical system. To identify the most pressing issues facing the system, in August 2024, GAO convened a discussion forum of a panel of experts and stakeholders. In this report, we examine forum participants' perspectives on factors that affect the system's ability to (1) build and maintain public trust, (2) meet the needs of its users, (3) sustain and modernize its data collection, and (4) engage in effective interagency coordination.

To prepare the discussion forum and this report, we conducted a systematic literature search and reviewed 51 studies concerning the

[6]89 Fed. Reg. 82453 (Oct. 11, 2024), codified at 5 C.F.R. part 1321.

[7]5 C.F.R. § 1321. Statistical Agencies and Units are defined under 44 U.S.C. § 3561(11) as executive branch agencies or organizational units whose primary activities involve the collection, compilation, processing, or analysis of information for statistical purposes. In accordance with 44 U.S.C. § 3562(a), the Director of OMB is authorized to recognize specific agencies or organizational units as statistical agencies and units. In this report, we use the phrase "statistical agencies" to encompass all recognized Statistical Agencies and Units.

[8]A statistical purpose is the description, estimation, or analysis of characteristics of whole or relevant groups within the economy, society, or national environment without identifying the individuals or organizations that comprise the groups. A nonstatistical purpose may include an administrative, regulatory, law enforcement, or other purpose that affects the rights, privileges, or benefits of a particular data provider. See 44 U.S.C. § 3561.

[9]Pub. L. No. 111-139, § 21, 124 Stat. 8, 29 (2010), codified at 31 U.S.C. § 712 note.

federal statistical system that met several criteria for inclusion.[10] With assistance from the National Academies of Sciences, Engineering, and Medicine (National Academies) we developed a list of experts and stakeholders and administered a questionnaire to them to inform forum topics and participant selection.

On August 21–22, 2024, we convened a group of 29 experts and stakeholders on the federal statistical system for a forum focused on challenges and opportunities faced by the system. The forum was planned and convened with the assistance of National Academies. The participants, selected to represent a range of experience and viewpoints, included representatives from the federal statistical system, other federal agencies, state and local government agencies, a non-U.S. national statistical office, an international organization, academic institutions, the private sector, and professional organizations. The 2-day forum was organized around six main topical sessions related to

- modernizing federal data collection and production;
- public use and accessibility of federal statistical products;
- resources, productivity, workforce, and efficiency in a modern federal statistical system;
- innovation in alternative data sources for federal statistics;
- enhancing data sharing across federal agencies; and
- public trust and objectivity in federal statistics.
- See appendix I for the forum agenda, appendix II for a list of the participants, and appendix III for details on our scope and methodology.

The forum was professionally recorded and transcribed. This report is a summary of the forum based on a thematic analysis of the discussion transcripts. The summary aims to capture the ideas and themes that emerged from the collective discussion of the participants and where appropriate supplemented by prepared written remarks from forum participants.

[10]Through a librarian-assisted search of databases using key terms, we first identified 76 relevant studies published within 5 years prior to and including September 2023. See appendix III for further information.

The forum was structured as guided roundtable discussions on each topic where four to six participants provided opening remarks and all participants were invited to openly comment on issues and respond to one another, although not all participants commented on all topics. Participants were given the opportunity to comment on a draft of this summary, and we included their feedback, as appropriate. After the forum, we also conducted three supplementary interviews to gain additional perspectives and to follow up on key themes discussed during the forum. We did not attempt to independently validate the statements expressed by participants.

Comments summarized in this report do not necessarily represent the views of all participants, the organizations with which they are affiliated, or GAO.

We conducted our work from October 2023 to September 2025 in accordance with all sections of GAO's Quality Assurance Framework that are relevant to our objectives. The framework requires that we plan and perform the engagement to obtain sufficient and appropriate evidence to meet our stated objectives and to discuss any limitations in our work. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

Background

The Federal Statistical System

The federal statistical system is a decentralized network that includes 16 statistical agencies and units in different federal departments or parent agencies.[11] A statistical agency is an entity within the executive branch whose activities predominantly are the collection, compilation, processing, or analysis of information for statistical purposes. The system spans diverse policy areas and provides critical input for program design, monitoring, and evaluation of federal programs (see table 1).

[11]According to OMB, the federal statistical system is a decentralized, interconnected network of 16 statistical agencies and units, 24 statistical officials (across 24 major cabinet agencies), approximately 100 additional federal statistical programs engaged in statistical activities, and several cross-system interagency and advisory bodies.

Table 1: Recognized Statistical Agencies and Units and Their Parent Agencies

Statistical Agency or Unit	Parent Agency	Focus Area
Bureau of Economic Analysis	Dept. of Commerce	Focuses on economic statistics to produce data on gross domestic product, personal income, and other economic indicators.
Bureau of Justice Statistics	Dept. of Justice	Collects, analyzes, and disseminates statistics on crime, criminal offenders, and the criminal justice system.
Bureau of Labor Statistics	Dept. of Labor	Provides information on the U.S. labor market and economic conditions by producing labor-related statistics, including the unemployment rate and inflation.
Bureau of Transportation Statistics	Dept. of Transportation	Gathers and analyzes data on transportation, including infrastructure, safety, and travel patterns.
Census Bureau	Dept. of Commerce	Conducts the decennial census and ongoing surveys, such as the American Community Survey, to provide demographic, economic, and housing information.
Center for Behavioral Health Statistics and Quality	Dept. of Health and Human Services	Operates within the Substance Abuse and Mental Health Services Administration to conduct national surveys tracking population-level behavioral health issues.
Economic Research Service	Dept. of Agriculture	Produces studies, economic analyses, and market assessments by examining factors such as commodity markets, farm income, and food assistance programs.
Energy Information Administration	Dept. of Energy	Specializes in energy-related statistics, including production, consumption, and distribution of energy resources.
Microeconomic Surveys	Federal Reserve Board	Operates within the Division of Research and Statistics to conduct research in a variety of areas, including consumer finances, financial markets, and general applied microeconomics.
National Agricultural Statistics Service	Dept. of Agriculture	Collects and disseminates agricultural statistics and analyses to provide information on the agricultural sector, crop production, and livestock, among other things.
National Animal Health Monitoring System	Dept. of Agriculture	Operates within the Animal and Plant Health Inspection Service to conduct national studies on the health and health management of U.S. domestic livestock, equine, aquaculture, and poultry populations.
National Center for Education Statistics	Dept. of Education	Focuses on education-related statistics, providing data on schools, educational attainment, and learning outcomes.
National Center for Health Statistics	Dept. of Health and Human Services	Operates within the Centers for Disease Control and Prevention to provide health-related data, including vital statistics, and health surveys.
National Center for Science and Engineering Statistics	National Science Foundation	Gathers and analyzes data on science and engineering research and development, education, and workforce.
Office of Research, Evaluation, and Statistics	Social Security Administration	Supports research, evaluation, and statistical analyses related to retirement, disability, and survivor benefits.
Statistics of Income Division	Dept. of Treasury	Collects, analyzes, and disseminates data on the income, taxes, and financial activities of individuals, businesses, and corporations.

Source: National Academies of Sciences, Engineering, and Medicine. | GAO-25-107124

Note: A statistical agency or unit is an entity in the executive branch whose activities are predominantly the collection, compilation, processing, or analysis of information for statistical purposes. The federal statistical system is a decentralized network of 16 statistical agencies and units (listed in table

1), 24 statistical officials (across 24 major cabinet agencies), approximately 100 additional federal statistical programs engaged in statistical activities, and several cross system interagency and advisory bodies.

While statistical agencies are typically embedded within larger parent agencies, they are expected to operate with a degree of autonomy to ensure the integrity, objectivity, and utility of the data they produce.[12] At the same time, these agencies contribute to the missions of their parent agencies by producing data that inform policy, evaluate programs, and support decision-making—demonstrating a dual role of statistical autonomy and institutional alignment with broader departmental goals.[13]

OMB's Role in Statistical Coordination and Oversight

OMB's Office of Information and Regulatory Affairs is charged with overseeing the use of information resources, including statistics, to improve the efficiency and effectiveness of governmental operations to serve agency missions, including burden reduction and service delivery to the public.[14] Within OMB, the Chief Statistician of the United States provides leadership on standards and interagency coordination through disseminating statistical policies and guidance.[15]

Specifically, OMB's statutory statistical responsibilities include the following:

- **Oversight and approval of data collection.** OMB is to review and approve proposed federal agency information collections that will be administered to 10 or more people, including minimizing

[12]The Information Quality Act, Pub. L. 106-554, 114 Stat. 2763A-153 to 2763A-154, (2000) "…requires OMB, as well as all other federal agencies, to maximize the objectivity, utility, and integrity of information, including statistical information, provided to the public." Office of Management and Budget, *Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units*, 79 FR 71610 (Dec. 2, 2014).

[13]According to the National Academies, it is imperative for statistical agencies to "provide objective, accurate, and timely information" while remaining "independent from political and other undue external influence in developing, producing, and disseminating statistics." See National Academies of Sciences, Engineering, and Medicine, *Principles and Practices for a Federal Statistical Agency: Eighth Edition* (Washington, D.C.: The National Academies Press, 2024), doi.org/10.17226/27934.

[14]44 U.S.C. § 3504(a)(1).

[15]OMB provides guidance to statistical units related to a variety of issues. See, for example, Office of Management and Budget, *Improving Statistical Activities through Interagency Collaboration*, M-15-15 (July 8, 2015); *Federal Data Strategy – A Framework for Consistency*, M-19-18 (June 4, 2019); and *Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance*, M-19-23 (July 10, 2019).

the information collection burden and maximizing the practical utility of information collected by or for the federal government.[16]

- **Guidance and standards.** OMB is to develop and oversee policies, principles, standards, and guidelines for federal information resources management, such as those for statistical activities, public access to data, and privacy and confidentiality.[17]
- **Coordination.** OMB is to coordinate the activities of the federal statistical system to ensure the integrity, objectivity, and utility of information, among other things.[18]
- **Oversight of budgets.** OMB is to ensure that statistical agencies' budget proposals are consistent with system-wide priorities for maintaining and improving the quality of federal statistics.[19]

As the coordinating body for the federal statistical system, OMB plays a central role in fostering collaboration and consistency across agencies, including through its leadership of the Interagency Council on Statistical Policy. The council is chaired by the Chief Statistician of the United States at OMB and brings together heads of major statistical programs to address cross-cutting issues. For example, the council has coordinated to align statistical agency practices with the Evidence Act and other reforms, which encourages statistical agencies to reevaluate existing approaches to data collection, product development, and service delivery. Specific

[16]44 U.S.C. § 3504(c), 5 C.F.R. § 1320.3(c).

[17]44 U.S.C. § 3504(a)(1). For example, OMB's Statistical Policy Directive No. 15 establishes uniform standards for collecting and reporting race and ethnicity data across all federal agencies, with the aim of ensuring consistency and comparability. See OMB, *Statistical Policy Directive No. 15: Standards for the Classification of Federal Data on Race and Ethnicity* (Mar. 29, 2024).

[18]44 U.S.C. § 3504(e)(1).

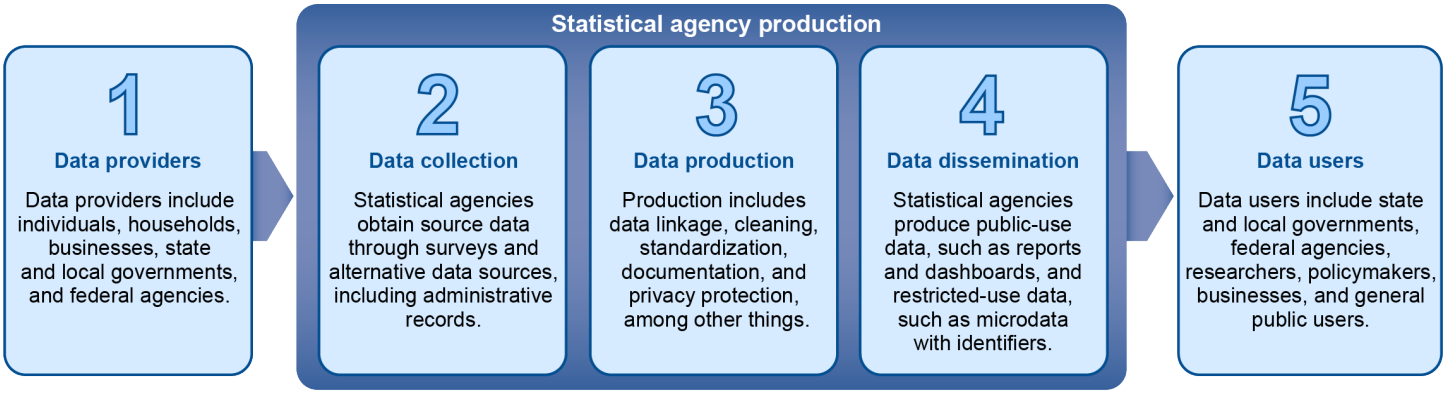
[19]44 U.S.C. § 3504(e)(2).

coordination efforts have included improving data access and advancing interagency data sharing and shared services.[20]

Data Providers and Users

Data providers and users play essential roles throughout the life cycle of federal statistical products—from initial data collection and processing to dissemination and application—ensuring that the federal statistical system both produces accurate information and delivers insights that inform decision-making and public understanding (see fig. 1).

Figure 1: Data Providers and Data Users in the Federal Statistical Product Life Cycle



Sources: GAO analysis of the Federal Statistical System Forum and National Academies of Sciences, Engineering, and Medicine. | GAO-25-107124

Data Providers

Data providers—including individuals, households, businesses, state and local governments, and federal agencies—supply critical raw data to the

[20]For example, the Interagency Council on Statistical Policy facilitates the implementation of the Standard Application Process, which provides a single, streamlined application portal for researchers to request access to confidential data across participating agencies. The Standard Application Process was developed in response to provisions in The Confidential Information Protection and Statistical Efficiency Act of 2018, which was enacted as part of title III of the Evidence Act. See Pub. L. No. 115-435, 132 Stat. 5529, 5554-56 (2019) (codified at 44 U.S.C. § 3583). See also Interagency Council on Statistical Policy, *SAP Governance Board Charter* (Washington, D.C.: Office of Management and Budget, December 2022). In addition, the Interagency Council on Statistical Policy supports the development of the National Secure Data Service-Demonstration Project, a key provision of The CHIPS and Science Act of 2022, which is designed to enhance secure data linkage and access for evidence-building across agencies. See Pub. L. No. 117-167, § 10375, 136 Stat. 1366, 1574 (2022). See also Interagency Council on Statistical Policy, *NSDS Subcommittee Charter* (Washington, D.C.: Office of Management and Budget, February 2024).

federal statistical system by responding to surveys and contributing to non-survey, alternative data sources.

Agencies are expected to collect data from providers using surveys designed specifically for statistical purposes. These surveys are intended to collect statistical data for the purpose of describing or making estimates on a wide range of topics. However, participation in federal surveys is often voluntary and has been declining over time, driven in part by concerns over privacy and confidentiality, as well as response burden.[21]

For example, according to the Census Bureau, the COVID-19 pandemic shutdowns caused major disruptions to the 2020 American Community Survey,[22] resulting in a decline in the response rate (to 71 percent).[23] The National Crime Victimization Survey led by the Bureau of Justice Statistics and the National Health Interview Survey led by the National Center for Health Statistics have also seen significant declines in response rates, with the former dropping from 80 percent in 2011 to 50 percent in 2021, and the latter from 70 percent to 50 percent in the same period. In addition to raising data collection costs, declining response

[21]In a study to understand potential barriers and motivators to participating in the decennial census, the Census Bureau found that concerns about the privacy and confidentiality of responses, and distrust of all levels of government were barriers to participating. See Kyley McGeeney et al., 2020 Census Barriers, Attitudes, and Motivators Study Survey Report: A New Design for the 21st Century, U.S. Census Bureau, Jan. 24, 2019. See also Jessica Holzberg, Jonathan Katz, and Mary Davis, *Measuring Respondents' Perceptions of Burden in the American Community Survey (ACS),* U.S. Census Bureau, Nov. 4, 2021.

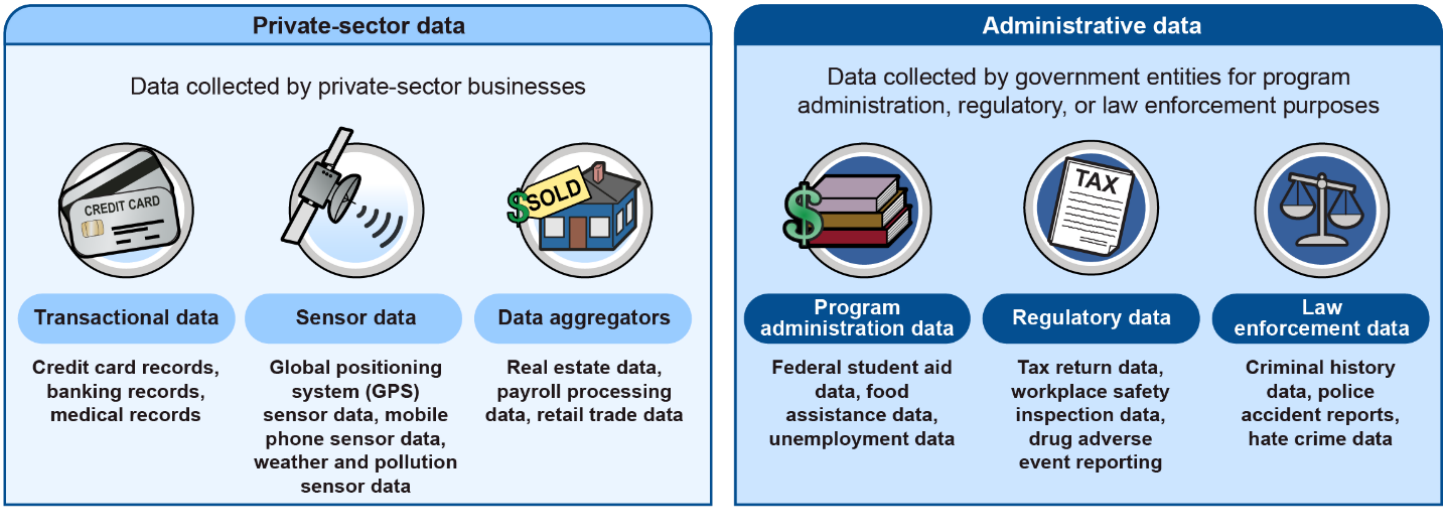
[22]The American Community Survey is an ongoing, nationwide survey conducted by the Census Bureau that collects detailed demographic, social, economic, and housing data at various levels of geography (e.g., national, state, county).

[23]While response rates for the survey have improved since the COVID-19 pandemic began in 2020—in 2023 the response rate for housing units was 85 percent—according to the Census Bureau, response rates are not yet back to prepandemic levels, which ranged from 92 percent in 2018 to 86 percent in 2019. See "Response Rates," American Community Survey, U.S. Census Bureau, accessed April 2, 2025, https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/.

rates have prompted concerns about data quality and representativeness.[24] In general, lower response rates increase the risk of less accurate estimates for statistical products.

Data providers also contribute to alternative data sources—that is, data collected for non-statistical purposes. These alternative data sources include transaction and consumer data from private-sector businesses and administrative records from federal programs, among other sources (see fig. 2). Through administrative records, where the information is provided to receive benefits or comply with the law, the government holds sensitive data about individuals, covering areas such as income, immigration, and health records.

Figure 2: Commonly Used Non-Survey Sources of Alternative Data



Sources: GAO analysis of the Federal Statistical System Forum and National Academies of Science, Engineering, and Medicine. | GAO-25-107124

Although alternative data sources were not originally created for statistical purposes, federal statistical agencies are increasingly exploring the blending of these data with traditional surveys to improve the timeliness,

[24]National Academies of Sciences, Engineering, and Medicine, *Nonresponse in Social Science Surveys: A Research Agenda* (Washington, D.C.: The National Academies Press, 2013), and *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy* (Washington, D.C.: The National Academies Press, 2017).

granularity, and relevance of national statistics.[25] This blended approach aims to address persistent challenges such as declining response rates and rising data collection costs.[26] For example, in response to initial COVID-19 pandemic shutdowns, the Bureau of Labor Statistics explored using private-sector transaction data, such as web-scraped data, to improve the accuracy and timeliness of price indices when in-person data collection methods (i.e., surveys) were unavailable.[27]

Data Users

Data users—including government entities, businesses, researchers, and the general public—rely on different types of data products, including aggregated data and microdata, to inform decisions, shape policies, and address community needs.

- **State and local governments** use federal statistical data—such as estimates on population, health, income, or housing—to inform local planning, resource allocation decisions, and to distribute state funding to localities. For example, they may use data from the Census Bureau to target policy interventions by region or demographic group within a state or locality.
- **Federal agencies** and programs use statistical data to evaluate program performance and allocate funding to various recipients. For example, agencies may integrate statistical data with

[25]While some statistical agencies have only recently started exploring blending administrative records with statistical data, some statistical agencies, such as the Bureau of Economic Analysis, have a long-standing history of using administrative data to produce key economic indicators. In addition, certain statistical programs rely exclusively on administrative data, such as the Statistics of Income. Specifically, data submitted to the Internal Revenue Service for tax administration serve as the foundation of its statistical products. These statistics are largely derived from statistical samples, enabling the collection of additional information from forms, schedules, and attachments, as well as the coding of data items to enhance their statistical utility.

[26]Studies from the National Academies suggest that blended data sources, when used responsibly, could mitigate the effect of declining survey response rates by providing more comprehensive and accurate datasets while potentially reducing data collection costs by leveraging already available data, reducing the need for expensive and time-consuming survey methods. See National Academies of Sciences, Engineering, and Medicine, *Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good* (Washington, D.C.: The National Academies Press, 2023); and *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*, 2017.

[27]National Academies of Sciences, Engineering, and Medicine, *Modernizing the Consumer Price Index for the 21st Century* (Washington, D.C.: The National Academies Press, 2023).

administrative records to address key policy questions.[28] In addition, several federal assistance programs use Census Bureau data—either in whole or in part—to guide funding allocations for areas such as health care, nutrition, highways, housing, school lunches, child care, and COVID-19 relief.

- **Researchers** and academics often depend on access to restricted-use microdata (i.e., data at the individual record level) for analysis in areas like public health and education.
- **Policymakers** may use statistical reports and aggregated findings to understand population needs, shape legislation, and conduct oversight. For example, social or economic statistics can influence decisions on funding formulas or the scope of a policy intervention.
- **Businesses** use federal statistical data for market analysis, risk modeling, and strategic planning.
- **General public users** may access summary statistics, dashboards, and public-use microdata to get statistical data relevant to their daily lives, including information on education, commuting, health, crime, and demographics, such as aging in their communities.

Data Privacy Protections

Federal statistical agencies operate under a legal framework designed to protect the confidentiality of the information they collect. Specifically, the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) enacted originally in 2002, and reauthorized and expanded in 2018, is a core statute that governs the use of data acquired for exclusively statistical purposes.[29] CIPSEA seeks to safeguard individually identifiable information collected for statistical purposes under a pledge of

[28]For example, the National Health Interview Survey has been linked with Medicaid and Medicare records to examine long-term health outcomes. See National Academies of Sciences, Engineering, and Medicine, *Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good*, 2023.

[29]CIPSEA of 2002, enacted as Title V of the E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899, 2962 (2002). This statute was reauthorized and expanded in 2018 by Title III of the Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 132 Stat. 5529, 5544 (2019), codified at 44 U.S.C. §§ 3501-3520.

confidentiality. To do so, it prohibits the disclosure of such data for any non-statistical purposes without the respondent's informed consent.[30]

CIPSEA authorizes limited sharing of business data among designated statistical agencies for statistical purposes[31] and requires non-statistical executive branch agencies to provide data requested from statistical agencies to the extent practicable, among other things.[32] It also includes direction for OMB to issue regulations to facilitate such sharing while establishing standards, to the extent possible, for complying with applicable laws requiring the protection and confidentiality of individually identifiable information.[33] In December 2022, OMB established a Standard Application Process for researchers, agencies, state and local governments, and other authorized users to apply to securely access confidential statistical data that were acquired for statistical purposes.[34]

Department and office-specific laws provide additional protections that impose confidentiality or use requirements and potential criminal penalties for unlawful disclosure (see table 2).

[30]See 44 U.S.C. 3572(c) outlining the limited circumstances under which data or information received exclusively for statistical purposes may be disclosed for any use other than an exclusively statistical purpose.

[31]The three designated statistical agencies are the Census Bureau, Bureau of Economic Analysis, and Bureau of Labor Statistics. See 44 U.S.C. § 3576(a). Business data are defined as operating and financial data and information about businesses, tax-exempt organizations, and government entities. See 44 U.S.C. § 3561(3).

[32]44 U.S.C. § 3582(a). This does not apply to any data asset that is subject to a statute that "prohibits the sharing or intended use of such asset in a manner as to leave no discretion on the issue." 44 U.S.C. § 3582(b).

[33]44 U.S.C. § 3582(c).

[34]CIPSEA of 2018 includes a requirement for OMB to establish a process (including a common application form) through which agencies and others may apply to access the data assets accessed or acquired by a statistical agency or unit produced as a result of statistical activities conducted for a statistical purpose. See Pub. L. No. 115-435, 132 Stat. 5529, 5554-56 (2019) (codified at 44 U.S.C. § 3583(a)). See also, 44 U.S.C. § 3561(6). According to OMB, "when assessing a proposed project's use of confidential data, statistical agencies and units must ensure that public trust and appropriate protections will be maintained. This means that the proposed use must be for an exclusive statistical purpose and conform to any statutory limitations that guide the use of those data." OMB, *Establishment of Standard Application Process Requirements on Recognized Statistical Agencies and Units*, M-23-04 (Dec. 8, 2022).

Table 2: Selection of Department/Office-Specific Statutory Provisions for Confidentiality of Data Collected for Statistical Purposes

Department/Office	Statutory Provision	Description
Department of Agriculture	7 U.S.C. § 2276	Includes limits on disclosure of certain information, including individually identifiable data collected for specific statistical purposes, and potential criminal penalties for violations.
Department of Commerce	13 U.S.C. §§ 9, 214	Includes a prohibition on the disclosure of any information that could identify respondents for any purpose other than the statistical purpose it was provided; potential criminal penalties for violations.
Institute of Education Sciences	20 U.S.C. § 9573	Prohibits use of individually identifiable information about students, their academic achievements, their families, and information with respect to individual schools for any purpose other than research, statistics, or evaluation, among other things. It also includes potential criminal penalties for violations.
Internal Revenue Service	26 U.S.C. §§ 6103, 7213	Generally restricts access to and disclosure of tax return information, ensuring taxpayer confidentiality with exceptions that include certain statistical uses by specified federal entities. Willful unauthorized disclosures may result in potential criminal penalties for violation.

Source: GAO analysis. | GAO-25-107124

In addition to statutory safeguards establishing legal foundations for protecting confidential information, statistical agencies also use privacy-enhancing technologies—such as synthetic data and differential privacy, among other disclosure avoidance methods—to further strengthen these protections throughout the data life cycle.[35] These technologies include the following:

- **Synthetic data.** Synthetic data are artificially generated datasets that replicate the statistical properties of real data without including actual respondent-level information. The technique

[35]According to the National Academies, synthetic data offers the potential to calculate aggregated estimates while reducing disclosure risks for individual records. For the various ways in which agencies have explored applying synthetic data technologies, see National Academies of Sciences, Engineering, and Medicine, *Toward a 21st Century National Data Infrastructure: Managing Privacy and Confidentiality Risks with Blended Data* (Washington, DC: The National Academies Press, 2024). However, note that some researchers have cautioned that synthetic data and differential privacy methods—while promising for confidentiality protection—can significantly limit the types of analyses that can be reliably performed, thereby potentially diminishing the value of federal statistical data. See V. Joseph Hotz, Christopher R. Bollinger, Tatiana Komarova, Charles F. Manski, Robert A. Moffitt, Denis Nekipelov, Aaron Sojourner, and Bruce D. Spencer, "Balancing Data Privacy and Usability in the Federal Statistical System," *Proceedings of the National Academy of Sciences* 119, no. 31 (2022).

allows users to conduct exploratory analysis and develop models without direct access to sensitive microdata. For example, the Census Bureau has developed synthetic datasets for the Survey of Income and Program Participation to support research use while reducing disclosure risk.[36]

- **Differential privacy.** Differential privacy introduces mathematically defined noise into statistical outputs that aims to preserve their value for statistical analysis, while limiting the ability to infer individual-level information. The Census Bureau implemented differential privacy for the 2020 Decennial Census through a disclosure avoidance system—the first large-scale use of this technique in a federal statistical product.[37]

The Federal Statistical System Faces Challenges in Building and Sustaining Public Trust, but Responsible Data Stewardship and New Reforms Offer Opportunities

Forum participants identified public trust as a cross-cutting issue that plays an important role in the modernization of the federal statistical system. Participants noted that the system faces growing challenges in building and maintaining public trust, particularly as it navigates emerging risks to the privacy and confidentiality of its data, as well as the autonomy of its statistical agencies. Participants discussed opportunities to reinforce public confidence in the system by embracing responsible data stewardship and public trust initiatives that promote transparency, safeguard privacy, and foster meaningful engagement across the statistical product life cycle.

[36]Jordan Stanley and Evan Totty, "Synthetic Data and Social Science Research: Accuracy Assessments and Practical Considerations from the SIPP Synthetic Beta," Working Paper Number CED-WP-2024-004, U.S. Census Bureau, 2024. Synthetic data are also used by agencies like the National Center for Health Statistics for public-use files. The term "disclosure risk" broadly means the probability that specified information about a particular data subject in a particular database and presumed private will be obtained by an unauthorized party and associated with the data subject.

[37]Disclosure avoidance methods are used to mitigate disclosure risk in public data products. These systems may incorporate traditional techniques (such as data swapping and suppression) alongside other tools like differential privacy. The intent of these systems is to ensure that released datasets cannot be reverse engineered to identify respondents.

Challenges in Building and Sustaining Public Trust in Federal Statistics

Risks to Privacy, Confidentiality, and Integrity

According to forum participants, public trust is fundamental to the federal statistical system, both from the perspective of data providers and data users. Data providers, including individuals, businesses, and organizations, must trust that their information, such as responses to surveys, will be kept confidential and used solely for statistical purposes. As one participant noted, when a statistical agency approaches individuals for data collection, the respondent must trust that the agency's confidentiality protections will safeguard their information, allowing them to report truthfully.

However, when trust in statistical agencies erodes due to concerns regarding disclosure risks or data leaks, respondents may refuse to participate in surveys and other federal data collection efforts. This may lead to a decline in survey response rates, and those who do respond may be less willing to provide truthful information, ultimately compromising the accuracy and reliability of federal statistical data.[38]

One Participant's View on Public Trust

"If there's a significant privacy violation then the loss of trust will lead to all of the future statistical product tables being unreliable…one demonstration of this is what I call the January 1st problem that various online services have, which is that a disproportionate number of people report their birthday to be January 1st."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

Similarly, data users rely on the integrity and autonomy of statistical agencies to provide accurate and unbiased information. According to one

[38]A recent report by the National Academies stated that trust among the public and data providers is critical to federal statistical agencies. Without the cooperation of data providers such as individuals, households, and businesses, federal statistical agencies would not be able to produce useful statistical information. See National Academies of Sciences, Engineering, and Medicine, *Principles and Practices for a Federal Statistical Agency: Eighth Edition*. Another study found that trust in statistical products and institutions is critical for ensuring the effectiveness of data collection efforts. The research highlights that trust in statistics may be influenced by factors such as transparency, credibility, and the public's experience with using federal data. See Jennifer H. Childs, Aleia C. Fobia, Ryan King, and Gerson Morales, "Trust and Credibility in the U.S. Federal Statistical System," *Survey Methods: Insights from the Field*, last modified February 22, 2019, https://surveyinsights.org/?p=10663.

participant, “there’s no real substitute” that provides the same level of rigor, consistency, and national representativeness as federal statistical products. Any perceived external influence or lack of transparency can erode confidence in federal statistical data and products, potentially leading to their diminished use in decision-making.[39] Thus, according to one participant “Trust in federal statistical agencies is extraordinarily important for both the collecting of statistical data and the reporting out of statistical information.”

Participants identified three challenges that statistical agencies face in maintaining public trust, as well as specific risks that may erode confidence in statistical products (see table 3).

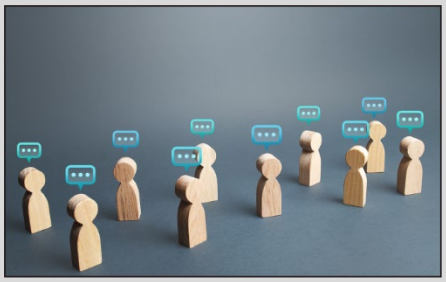
Table 3: Reported Challenges Affecting Public Perceptions of Privacy, Confidentiality, and Integrity of Federal Statistical Agencies

Challenge	Description
Evolving data ecosystems	Evolving data ecosystems, such as expanding users and the use of administrative records, can increase perceived risks to confidentiality.
Use of statistical products outside of intended scope	Using statistical products outside their intended scope can erode public trust.
Role distinction	The public may misunderstand the distinct roles between statistical activities and other policy functions.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

[39]A recent report by the National Academies emphasizes that transparency is essential to building and maintaining public trust and concludes that agencies should demonstrate independence from external influences. See National Academies of Sciences, Engineering, and Medicine, *Transparency in Statistical Information for the National Center for Science and Engineering Statistics and All Federal Statistical Agencies* (Washington, D.C.: The National Academies Press, 2022).

Gauging Public Trust in Statistical Activities



Some federal statistical agencies assess public trust in the federal statistical system through surveys and feedback tools designed to gauge perceptions of credibility, transparency, and data stewardship. The Census Bureau includes trust-related questions in several of its efforts, including the Household Pulse Survey, which began collecting insights related to public trust. Additionally, the Census Barriers, Attitudes, and Motivators Study explores public perceptions of federal data collection, including concerns about privacy, data use, and credibility. Similarly, the Bureau of Labor Statistics introduced a customer satisfaction survey on its public website in April 2024. One of the key questions asks users whether they agree or disagree that the Bureau of Labor Statistics is a trusted source of information and to provide direct feedback on the agency's reputation for reliability.

Source: Andrii Yalanskyi/stock.adobe.com (image); GAO analysis of agency documentation and forum follow-up; Kylee McGeeney et al., 2020 Census Barriers, Attitudes, and Motivators Study Survey Report (U.S. Census Bureau, 2019). | GAO-25-107124

Evolving data ecosystems. Participants warned that a breach of confidentiality could affect how the public perceives the statistical system, and that changes in data ecosystems—such as modernization efforts to expand access to statistical data to a range of data users and to create products from data that is blended from a variety of sources—make efforts to implement privacy and confidentiality safeguards more urgent.[40] In addition, expanded use of alternative data sources, such as administrative and proprietary data, to supplement statistical data also raises privacy concerns, as these data often lack the same confidentiality protections and governance frameworks as traditional federal statistical data.

While federal surveys typically include robust consent procedures that clarify how data will be used, such transparency and legal protections are not always present for administrative or proprietary data. In some cases, there is no overarching legal framework ensuring consistent privacy protections for these data when used for statistical purposes, although some safeguards may exist depending on the source or

[40]Concerns regarding data privacy and confidentiality are reflected in our larger body of work that examines how federal agencies handle personal data and highlight key challenges in safeguarding information that can undermine public trust. For key reports on this topic, see GAO, Protecting Personal Privacy, accessed May 22, 2025. See also GAO, *High-Risk Series: Urgent Action Needed to Address Critical Cybersecurity Challenges Facing the Nation*, GAO-24-107231 (Washington, D.C.: June 13, 2024). As of April 2025, about 104 of our 260 privacy-related recommendations remained unimplemented. Many agencies lacked dedicated privacy leadership or had not fully integrated privacy into risk strategies.

agency.[41] Linking traditional survey data with less-protected sources may heighten data providers' concerns about misuse or reidentification, leading to skepticism and reduced willingness to share information.

Use of statistical products outside of intended scope. Using statistical products outside of their intended scope can also erode trust in the federal statistical system. For example, a forum participant raised concerns about policymakers' use of the National Risk Index, the Federal Emergency Management Agency's aggregation of data from the federal statistical system.[42] The Federal Emergency Management Agency developed the National Risk Index using data from the Census Bureau, among other data sources, to identify areas at risk from natural hazards, such as floods, wildfires, and hurricanes.

However, policymakers have used the National Risk Index as a tool to inform mitigation planning and data-driven decision making aimed at disaster preparedness and resilience across localities. The forum participant discussed how a data user told them that the National Risk Index quickly took on outsized importance relative to local data and first-hand observations. Further, the data user found that relying solely on the National Risk Index to allocate funding to localities may be flawed, as the index was not originally designed for that purpose—potentially leading to funding decisions that misalign with actual local needs.

Role distinction. Participants discussed how a general lack of trust in the U.S. government can make it difficult for statistical agencies to maintain public trust in the federal statistical system, particularly when the public may not differentiate statistical activities from other government functions. One forum participant highlighted a key finding from a recurring survey of international, high-income countries, which found that trust in government

[41]The collection or use of personal information by the federal government is governed primarily by two laws: the Privacy Act of 1974, as amended, and the privacy provisions of the E-Government Act of 2002. See Pub. L. 93-579, 88 Stat. 1896 (1974), codified as amended at 5 U.S.C. § 552a, and Pub. L. 107-347, 116 Stat. 2899 (2002), codified at 44 U.S.C. § 101. However, there is no overarching federal privacy law that governs the collection and sale of personal information among private-sector companies. In 2013, we recommended that Congress consider strengthening the consumer privacy framework to reflect the effects of changes in technology and the increased market for consumer information. As of February 2025, Congress has not enacted comprehensive privacy legislation that would address this matter. See GAO, *Information Resellers: Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace*, GAO-13-663 (Washington, D.C.: Sept. 25, 2013).

[42]The National Risk Index, Federal Emergency Management Agency, accessed March 27, 2025, https://hazards.fema.gov/nri/.

statistics is closely associated with levels of trust in government.[43] Another participant cited an annual study of public trust in government, which estimated that 16 percent of Americans trusted the government "just about always" or "most of the time"—the lowest in 7 decades of polling.[44]

Participants suggested that public outreach and clearly communicating the difference between statistical activities and other government functions could clarify statistical agencies' objectivity and lack of outside influence. This distinction helps assure the public that data are used solely for statistical activities and will not be misused. According to participants, public misunderstanding in these roles could raise concerns about how the data are used, potentially affecting perceptions of the system's reliability and integrity.

Statistical Agency Autonomy Within Parent Agencies

According to participants, the autonomy of statistical agencies within their parent agencies plays a critical role in building and sustaining public trust. Agencies may struggle to ensure the transparency, accountability, and integrity of federal statistical data without sufficient autonomy over how data are collected, processed, and disseminated, which may ultimately affect public confidence in the system. Participants noted that the placement of statistical agencies within larger parent agencies may limit their autonomy in three ways: misalignment of goals, limited interaction with congressional policymakers, and lack of direct budget input (see table 4).

[43]Organization for Economic Cooperation and Development, *OECD Survey on Drivers of Trust in Public Institutions – 2024 Results: Building Trust in a Complex Policy Environment* (OECD Publishing, Paris, 2024), doi.org/10.1787/9a20554b-en.

[44]Pew Research Center, *Public Trust in Government: 1958-2024* (Pew Research Center, June 24, 2024), https://www.pewresearch.org/politics/2024/06/24/public-trust-in-government-1958-2024/.

Table 4: Reported Challenges Affecting Statistical Autonomy Within Parent Agencies of the Federal Statistical System

Challenge	Description
Misalignment of goals	Parent agencies may have policy, operational, or regulatory missions that do not include statistical priorities.
Limited interaction with congressional policymakers	Statistical agencies may have insufficient mechanisms to communicate directly with congressional stakeholders.
Lack of direct budget input	Statistical agencies may not have had a mechanism to directly participate with Congress in the annual budget formulation process.[a]

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

[a]Following the Federal Statistical System Forum on August 21-22, 2024, the Trust Regulation, effective December 10, 2024, requires statistical and parent agencies to jointly coordinate to develop a separate budget request for the statistical agency, submitted as part of the parent agency's budget to the Office of Management and Budget (OMB). The regulation also ensures the statistical agency head can participate in related OMB discussions. See 5 C.F.R. § 1321.4(g)(1-2).

Misalignment of goals. Participants noted that it is sometimes hard to determine how a statistical agency's goals align with its parent agency's goals. While statistical agencies are expected to function in an environment that is autonomous from other administrative, regulatory, law enforcement, or policy-making activities within their parent agency to ensure objectivity, they are also expected to collaborate with their parent agencies to enhance the relevance and usefulness of statistical products.[45] In addition, statistical agencies are expected to support priorities to maintain and improve the quality of federal statistics as directed by OMB.[46] However, according to participants, the extent to which these statistical priorities are incorporated into the broader goals and priorities of parent agencies is not always clear, which may limit the effectiveness of such collaboration and harm public trust.

For example, one participant cited a 2007 advocacy report that highlighted the lack of clarity regarding how the Centers for Disease Control and Prevention's public health objectives—emphasizing public health interventions, disease prevention, and preparedness—aligned with the National Center for Health Statistics mission to produce health statistics.[47] The report stated that this lack of clarity raised concerns about whether the National Center for Health Statistics could maintain its

[45]Office of Management and Budget, *Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units*, 2014.

[46]Office of Management and Budget, *Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units*, 2014.

[47]Population Association of America, "Safeguarding Vital and Health Statistics: Prepared for the Committee on Population Statistics," Population Association of America, October 2007.

independence as a statistical agency, given its location within the Centers for Disease Control and Prevention. According to forum participants, a lack of alignment and institutional support between a parent agency and its statistical agency can potentially limit the latter's autonomy and ability to maintain the confidentiality of sensitive data, undercut budget and staffing support, and constrain professional decisions on data production and dissemination.

Limited interaction with congressional policymakers. Participants discussed concerns regarding the ability of statistical agencies to have greater autonomy in the legislative process, whether individually or as a system. One participant noted that congressional hearings on oversight and appropriations often do not include federal statistical representatives. The parent agency generally designates a cabinet secretary or other senior official who may choose to prioritize speaking to the needs of agencies and programs under the parent organization's jurisdiction, other than the statistical agency. One participant concluded that this may have "impeded the ability of [statistical agencies] to communicate openly about their work and their desired objectives."[48] For example, participants noted that some smaller statistical agencies, especially those with limited visibility within their departments, often lack both the visibility and opportunity to share their insights on user needs directly with policymakers.

Participants also reported challenges with understanding the needs of policymakers. Participants said that since direct communication with Congress typically occurs through formal mechanisms—such as hearings or briefings—and statistical agencies are rarely invited to participate, they have limited opportunities to build reciprocal relationships with policymakers. As a result, they may struggle to demonstrate the value of statistical products or to gain insight into what public officials and their constituents need from the statistical system.

[48]The four fundamental responsibilities of statistical agencies: (1) produce and disseminate relevant and timely statistical information; (2) conduct credible and accurate statistical activities; (3) conduct objective statistical activities; and (4) protect the trust of information providers by ensuring the confidentiality of their responses and the exclusive statistical use of the responses. 44 U.S.C. § 3563(a)(1).

Statistical Agencies and Declining Access to Resources

In its 2024 report, The Nation's Data at Risk, the American Statistical Association highlighted that while statistical agencies are generally fulfilling their legal responsibilities, their ability to innovate and meet growing data demands is constrained due, in part, to declining access to resources. The report found that over the past 15 years, funding for most statistical agencies has declined 14 percent in purchasing power, which the report attributes to their limited ability to secure authority for multiyear funding and communicate the importance of their work directly to policymakers. These financial constraints have reduced their capacity to innovate, respond to emerging data needs, and even maintain existing operations, which may affect the overall utility and integrity of the data statistical agencies collect and produce.

Source: Alexkava/stock.adobe.com (image); GAO analysis of American Statistical Association, The Nation's Data at Risk: A Call to Secure and Modernize Federal Statistical Infrastructure (Alexandria, VA: American Statistical Association, July 9, 2024). | GAO-25-107124

Lack of direct budget input. At the time of the forum in August 2024, participants said that some statistical agencies did not have a formal mechanism to directly participate in the annual budget request and formulation process with their parent agencies for submission to OMB.[49] This has posed a challenge for statistical agencies embedded within larger departments. In addition, participants noted that some statistical agencies receive funding through their parent agencies' appropriations rather than having a dedicated line-item appropriation, which could result in statistical priorities being deprioritized in favor of broader departmental objectives. This means that there was not a dedicated way for policymakers to focus on statistical agencies and their specific resource needs.

Effective in December 2024, subsequent to the forum, the Trust Regulation directs joint coordination between the statistical and parent agencies in the development of a budget request for the statistical agency. This request is a separate part of the parent agency's annual budget request. Parent agencies are also required to ensure that the statistical agency head can participate in discussions with OMB related to that request.[50] However, the practical effects of the regulation remain to be seen, as its implementation is still in the early stages.

Opportunities to Build and Sustain Public Trust in Federal Statistics

[49]The Director of OMB ensures that agency budget proposals are consistent with system-wide priorities for maintaining and improving the quality of federal statistics. 44 U.S.C. § 3504(e)(2). In addition, the Interagency Council on Statistical Policy, led by the Chief Statistician, coordinates on cross-cutting statistical issues, including funding priorities. 44 U.S.C. § 3504(e)(8).

[50]5 C.F.R. § 1321.4(g)(1-2).

Balancing Statistical Utility with Data Privacy

According to participants, strengthening public trust in federal statistics depends on statistical agencies balancing the production of high-quality, accessible data products that meet public and policy needs with data privacy. Forum participants urged statistical agencies to prioritize product development based on the needs of data users, rather than focusing on routine processes or producing outputs without clear value or utility.[51]

One Participant's View on Statistical Product Utility

"We're here to serve people, so that they can make informed decisions about their lives, and their businesses, about the things they care about."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

To do this, participants stated that statistical agencies need to better understand and anticipate the various opportunities that could enhance the system's ability to balance statistical utility with privacy and confidentiality (see table 5).

Table 5: Reported Opportunities for Statistical Agencies to Balance Statistical Utility with Data Privacy

Opportunity	Description
Understand data interests	Statistical agencies could understand the privacy expectations of data providers and utility needs of data users.
Support secure data enclaves	Statistical agencies could support access to sensitive data within secure environments.
Incorporate feedback	Statistical agencies could engage with data providers and users to refine utility and privacy trade-offs.
Increase transparency	Statistical agencies could clearly communicate data practices, limitations, and protections, which could help statistical agencies increase transparency and build public trust.
Implement privacy-enhancing technologies	Statistical agencies could implement advanced privacy techniques to safeguard data while maintaining usability.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

[51]A 2023 report by the National Academies of Sciences, Engineering, and Medicine stresses that federal statistical efforts must focus on producing information that is useful, timely, and relevant to data users—especially amid declining survey response rates and increasing data collection costs associated with traditional sampled surveys. The report emphasizes that the success of a modernized data infrastructure for the federal statistical system relies on engaging stakeholders, piloting blended data approaches that protect privacy, and using clear criteria for data prioritization. The report highlights that expanding the utility of data must be ethically grounded, noting it is both "ethically necessary and technically possible" to preserve privacy and fulfill confidentiality commitments while increasing the statistical usefulness of diverse data sources. See National Academies of Sciences, Engineering, and Medicine, *Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good*, 2023.

Understand data interests. Participants discussed how the utility concerns of data users may differ from the privacy concerns of data providers, and understanding these differences could assist the federal statistical system's ability to pinpoint strategies to enhance public trust. For example, data users regularly request more granular data. However, more granularity divides data into smaller subgroups which may risk the reidentification of respondents.

One Participant's View on Understanding Data Interests

"It is no longer simply a balance between the accuracy of the statistics and the confidentiality or privacy protection of those statistics. Rather, there is a third component to consider, which is the quantity and granularity of the statistics you release. We call this availability, and statistical agencies need to balance across all three dimensions of this triple tradeoff (accuracy, availability, and confidentiality)."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

Product Utility and Data Privacy: Census Bureau's Disclosure Avoidance System

The Census Bureau's differential privacy simulation underscored the challenge of balancing data privacy with product utility. In 2018, the Census Bureau conducted an experiment to simulate database reconstruction that demonstrated how published data could be used to recreate confidential individual-level information, risking the reidentification of tens of millions of respondents. In response, the Bureau implemented a new Disclosure Avoidance System for the 2020 Census, incorporating statistical noise, or data uncertainty, through differential privacy to enhance confidentiality. While this approach improves privacy protection, some argue it may reduce the accuracy of demographic and geographic data, particularly at local levels. The agency has continued refining the system and plans to use differential privacy for future products. Following our recommendation, the agency acknowledged the need for a reliable schedule and, as of June 2024, provided updated timelines for disclosure avoidance activities—improving its ability to plan and track progress on protecting respondent data.

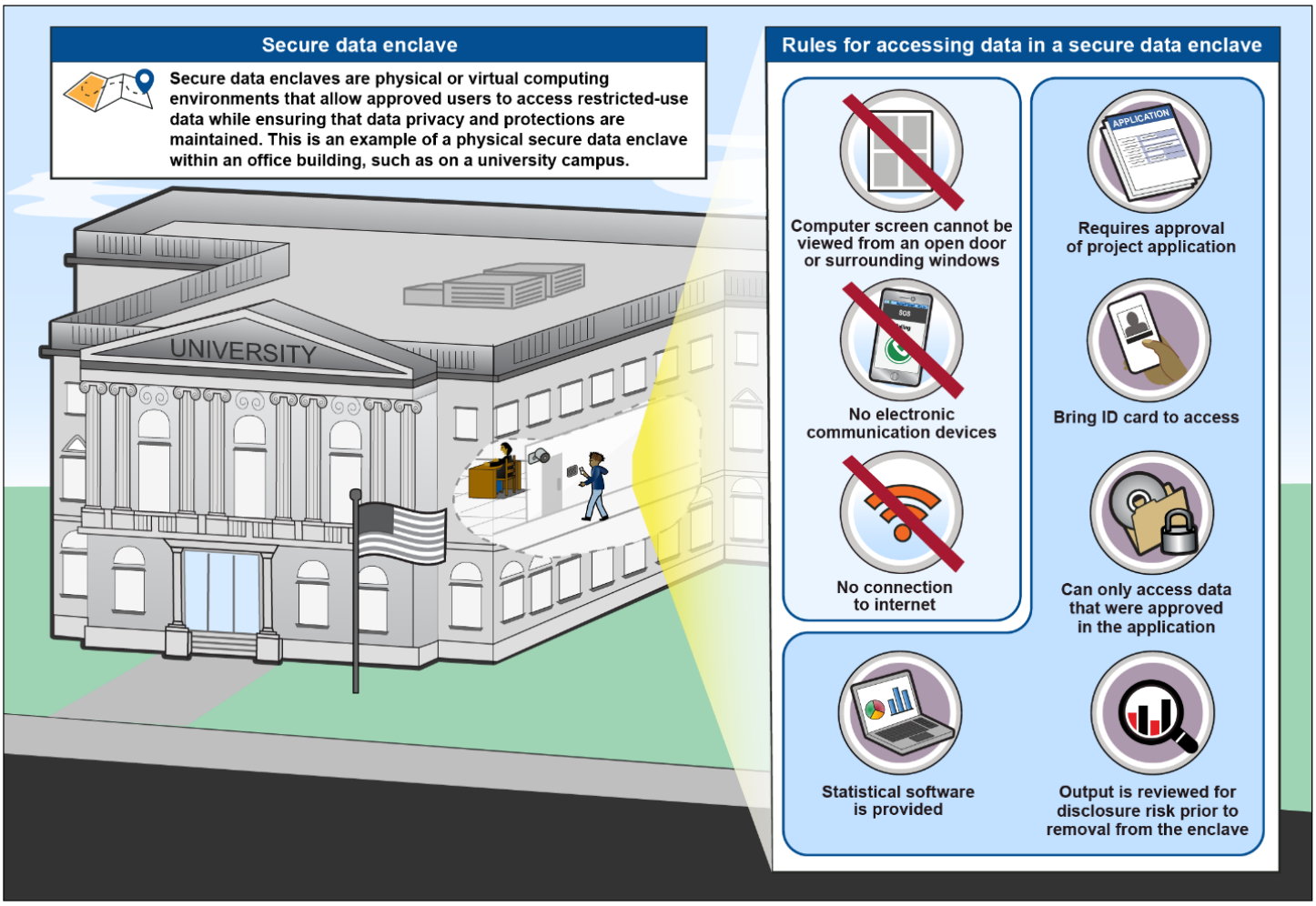
Source: GAO, 2020 Census: Bureau Released Apportionment and Redistricting Data, but Needs to Finalize Plans for Future Data Products, GAO-22-105324 (Washington, D.C.: Mar. 15, 2022) and John M. Abowd et al., "A Simulated Reconstruction and Reidentification Attack on the 2010 US Census," Harvard Data Science Review, vol. 7, no. 3 (2025), accessed September 9, 2025, doi.org/10.1162/99608f92.4a1ebf70. | GAO-25-107124

There are privacy protections and privacy-enhancing technologies that agencies adopt to mitigate these risks. However, one forum participant said some data users believe that statistical agencies are overly degrading the usefulness of the statistics they release as they seek to keep pace with the ever-increasing disclosure risk of those statistics.[52] Stronger confidentiality protections inevitably reduce the usability of statistical products, with greater protection generally leading to greater degradation. Yet participants also emphasized that enhancing the usefulness of statistical data is not necessarily at odds with ensuring confidentiality. Balancing both needs is critical because breaches of confidentiality, such as unauthorized disclosures, and risks of reidentification of individuals from publicly released statistical data by third parties can violate privacy and undermine public trust. Federal statistical products often rely on the goodwill of the public to respond truthfully to surveys and for data providers to provide accurate information. Combined with already decreasing survey response rates, a decline in public trust could reduce the precision and accuracy of statistical products.

Support secure data enclaves. To strike the right balance between product utility and the goals of confidentiality, participants stressed the importance of using risk mitigation strategies during data collection, analysis, and blending, such as supporting physical and virtual data enclaves to access restricted-use data. Secure data enclaves are restricted-access computing environments that allow approved users, such as researchers at universities or government agencies, to work with confidential or sensitive datasets while ensuring that data privacy and confidentiality protections are maintained. These enclaves have strict access controls, physical and cybersecurity protections, and review processes to prevent the unauthorized use of, or reidentification of individuals within, confidential data. Secure data enclaves are sometimes accessed in a specific building or facility, as shown in figure 3.

[52]For a discussion on the "triple tradeoff" in federal statistics (i.e., the tension between the priorities of accuracy, availability, and confidentiality of statistical products), see M.B. Hawes et al., "Toward a Principled Framework for Disclosure Avoidance," *Harvard Data Science Review* (2025), doi.org/10.1162/99608f92.db29c137. For example, statistical agencies must balance these competing priorities, as increasing availability and accuracy often raises disclosure risk, while stronger confidentiality protections can degrade data usefulness.

Figure 3: Characteristics of a Physical Secure Data Enclave



Sources: GAO illustration and analysis of the Federal Statistical System Forum, National Center for Health Statistics, Inter-university Consortium for Political and Social Research. | GAO-25-107124

Participants provided three illustrative examples of secure data enclaves, both physical and virtual. For example, Federal Statistical Research Data Centers are secure facilities managed by the Census Bureau in partnership with various federal statistical agencies and research institutions. They provide access to restricted-use microdata from numerous government surveys and censuses.[53] One participant also

[53]There are 35 centers across the United States, primarily located at academic institutions and Federal Reserve Banks. According to Census officials, the Federal Statistical Research Data Centers pivoted to a virtual access model at the start of the COVID-19 pandemic in 2020, which the agency hopes to continue to support to broaden its user base to state and local governments.

discussed the Administrative Data Research Facility, which is cloud-based platform designed to facilitate the access and analysis of confidential microdata.[54] In a virtual environment, the platform enables government agencies to link their data with other states and agencies. Participants also discussed balancing data utility and confidentiality in the National Secure Data Service, a proposed virtual secure data enclave that seeks to streamline access to restricted-use data across statistical agencies.

Incorporate feedback. Participants stressed the need to incorporate data provider and user perspectives on balancing utility and data protection concerns throughout the product cycle. For example, stakeholder feedback on acceptable risks and the usefulness of statistical products can help inform decision-makers about tradeoffs. As one participant told us, this is necessary, given the unavoidable limitations of risk mitigation.

One Participant's View on Incorporating Feedback

"Any data release, blended or not blended, that offers non-trivial usefulness, introduces disclosure risks. There is no method that guarantees zero disclosure risk unless it doesn't use the data. … We need policymakers, data owners, researchers to provide input on … not only the level of risk they're willing to accept, but also the level of usefulness of a particular data release."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

Participants recommended involving privacy experts and advocates in early discussions on what to collect and what information to publish. Retrofitting products to meet confidentiality standards often results in greater losses of utility, as privacy protections are applied after the fact rather than being integrated into the design from the outset. Another participant also suggested giving data providers "a seat at the table" in decisions on how their data will be used.

Increase transparency. Participants said that greater balance between product utility, privacy, and confidentiality could be achieved through more transparency on the release and access of statistical products, as

[54]The Coleridge Initiative, a nonprofit organization working to improve the use of data for decision-making and data literacy, developed the Administrative Data Research Facility under the guidance of the Census Bureau.

well as clear communication on how agencies plan to use data provider information and ensure provider confidentiality. A participant discussed how their agency has communicated with data providers about how data from their communities are used in statistical products and research to foster transparency and trust. Other participants noted that clarifying the difference between publicly available data versus restricted access data would be beneficial, and statistical agencies should communicate how they ensure privacy and confidentiality for both types of data.

Implement privacy-enhancing technologies. Participants raised the need to focus on privacy-enhancing technologies, such as implementing synthetic data and differential privacy techniques, to mitigate public concerns over privacy and confidentiality. As one participant noted, while these technologies are not new to the system, only recently have they been included in discussions regarding ways to balance public trust and utility of statistical products, particularly with regard to the sharing and access of statistical data. In addition to synthetic data and differential privacy methods, participants noted that statistical agencies are exploring the feasibility of secure multiparty computation and secure query systems.[55]

An International Example of Privacy-Enhancing Technologies

One participant provided an international example of a model for implementing privacy-enhancing technologies across statistical agencies and programs. According to the participant, Finland supports various statistical agencies by enabling secure, efficient data sharing and management across government departments, including the secure integration of survey data with administrative data. Finland's infrastructure employs several privacy-enhancing technologies, including the implementation of synthetic data, differential privacy, and masking.

Source: Thitichaya/stock.adobe.com (image); GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Strengthening Autonomy Through the Trust Regulation

The Trust Regulation aims to clarify the roles and responsibilities of statistical agencies, as well as how parent agencies should support them.[56] The regulation, which was issued in October 2024, specifies how parent agencies should support statistical agencies in carrying out their fundamental responsibilities. At the time of the forum in August 2024, participants expected that the implementation of the Trust Regulation, as proposed in August 2023, would build public trust in federal statistics by establishing a framework for statistical autonomy (see table 6).

[55]Secure multiparty computation is a method for researchers to analyze data from multiple data owners without revealing the sensitive input data. Secure query systems are a method by which users request predefined output statistics about specific records from statistical agencies without needing to access sensitive input data.

[56]OMB's Trust Regulation, officially referred to as the *Fundamental Responsibilities of Statistical Agencies and Units* rule, is codified at 5 C.F.R. part 1321. This implements the statutory mandate from the Evidence Act, Pub. L. No. 115-435, 132 Stat. 5529, 5547 (2019), codified at 44 U.S.C. § 3563(c).

Table 6: Selected Approaches in the Trust Regulation to Enhance Federal Statistical Agency Autonomy

Topic	Provision	How Autonomy Is Strengthened
Coordination	5 C.F.R. § 1321.4(b-c)	Requires regular communication between the parent and statistical agencies. Ensures parent agency policies support statistical agencies in meeting their requirements to maintain accurate information and produce relevant and timely statistical products.
Branding	5 C.F.R. § 1321.4(e-f)	Statistical agencies are to maintain distinct branding for specified public-facing communications, including stand-alone websites and statistical products. Parent agencies, in turn, must provide the necessary resources to maintain the statistical agency's website, as well as the authority and autonomy to allow the statistical agency to manage and update its website.
Budget Proposals	5 C.F.R. § 1321.4(g)(1-2)	Statistical and parent agencies must jointly coordinate to develop a distinct budget request for the statistical agency, which is submitted as a separate component of the parent agency's annual budget submission to the Office of Management and Budget (OMB). Parent agencies must also ensure that the statistical agency head can participate in OMB discussions related to that request.
Resources & Capacity	5 C.F.R. § 1321.4(h)	With statistical agencies, parent agencies must jointly develop options for addressing identified capacity needs of statistical agencies and make the necessary resources available to the extent practicable. The statistical agency is to notify OMB if it has insufficient capacity.
Publishing Statistics	5 C.F.R. § 1321.5	Statistical agencies are to control what, when, and how data are released, consult broadly with external data users, and ensure transparency with public release schedules with the support of the parent agency.
Maintaining Objectivity	5 C.F.R. § 1321.7	Parent agencies are to take certain steps to ensure the statistical unit's statistical work is objective. The statistical agency must ensure equitable data access by data users, and authority to grant access to confidential statistical data.

Source: GAO analysis. | GAO-25-107124

Note: Issued in October 2024, OMB's Trust Regulation, officially referred to as the Fundamental Responsibilities of Statistical Agencies and Units rule, is codified at 5 C.F.R. part 1321, implementing the statutory mandate under 44 U.S.C. § 3563.

According to participants, the Trust Regulation can support public confidence in the integrity and objectivity of federal statistics. The regulation would assist statistical agencies in branding themselves as well as conducting outreach by reinforcing their commitment to autonomy, objectivity, and data confidentiality. This can include educational campaigns, transparency initiatives, and clearer messaging on the role of federal statistics in decision-making.

The Federal Statistical System Faces Challenges in Meeting Diverse User Needs That May Be Mitigated by Expanding Data Accessibility

The federal statistical system serves a diverse user base, from highly technical researchers to non-technical data users. According to participants, these users rely on federal statistical data for various applications, yet the federal statistical system faces challenges in meeting user needs. In addition, participants said that the system also faces challenges in offering appropriate support and tools tailored to different user groups. To address these challenges, participants suggested that statistical agencies can expand data access through infrastructure and service-based initiatives, while enhancing data capacity of its users by providing training, guidance, and outreach.

Challenges in Meeting Diverse Data and Accessibility Needs of Users

Meeting Diverse Data Needs

Forum participants identified challenges in meeting data needs among users (see table 7).[57]

Table 7: Reported Challenges for the Federal Statistical System in Meeting Data Needs of Users

Challenges	Description
Gaps in granularity	Granular or disaggregated data may not be available for small geographic levels or for specific groups.
Gaps in timeliness	Statistical products may be delayed or may fail to keep up with emerging user needs.
Gaps in relevance	Existing statistical products may not always meet the evolving or rapidly changing needs of data users.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

[57]These challenges in meeting data needs are also selected key dimensions of data quality. See Federal Committee on Statistical Methodology, *A Framework for Data Quality, FCSM 20-04* (September 2020), https://www.fcsm.gov/assets/files/docs/FCSM.20.04_A_Framework_for_Data_Quality.pdf.

Constraints in Providing Granular Data

Providing granular data is often limited by privacy concerns and resource constraints. Participants told us that statistical agencies must make tradeoffs between the specificity of the data they release and the risk of a breach of confidentiality. In addition, providing granular data requires significant resources, as it often involves larger sample sizes and enhanced data processing. Multiple participants noted that resource constraints limit the federal statistical system's ability to provide users access to granular and relevant data. As one participant said, "Budgetary constraints impact our ability to innovate and to meet the growing demand for granular and interconnected data that's essential for evidence-based policy making."

Source: Dilok/stock.adobe.com (image), GAO analysis of Federal Statistical System Forum, and National Academies of Sciences, Engineering, and Medicine report. | GAO-25-107124

Gaps in granularity. According to participants, policymakers, state agencies, and local officials often require highly granular data—statistics at small geographic levels (e.g., city or neighborhood)—to target resources and craft policies for specific communities. Granular data can reveal local characteristics that state or national averages might conceal, enabling more effective policy interventions. According to participants, users seek more granular data, but federal statistical products often prioritize estimates for larger geographic levels, such as the national, state, and metropolitan area unemployment rates, due to their being broadly useful in supporting policy decisions for the nation and minimizing disclosure risk.[58]

Gaps in the availability of granular data for certain geographic areas present additional challenges in meeting this need. For example, a federal agency official we interviewed discussed how even after agencies blend data from multiple sources—that is, combining previously collected data sources—population, economic, and social estimates relevant to Native American populations or geography are often not represented in public products.[59] As a result, users do not always have sufficient data to support decision-making for policies and programs relevant to these groups and areas.[60]

Gaps in timeliness. Participants discussed user needs for timely data, including more frequent and real-time data updates. However, these

[58]While granular data are available for selected datasets, such as the Bureau of Labor Statistics' Local Area Unemployment Statistics that provides monthly estimates of unemployment for counties and some cities and towns, granular data are not available for other datasets, especially those involving sample surveys due to small sample sizes for granular geographic areas.

[59]Mohr, Caryn, "Beyond the gaps: Taking on data needs in Indian Country," Federal Reserve Bank of Minneapolis (Jan. 18, 2023), https://www.minneapolisfed.org/article/2023/beyond-the-gaps-taking-on-data-needs-in-indian-country.

[60]In representative surveys, it can be challenging to obtain sufficient responses from small population groups to report reliable estimates. In 2024, we reported that cost, geographic, technical, and other issues contribute to significant data gaps for U.S. territories. We recommended that OMB develop a coordinated, government-wide approach to identify ways to address these gaps. See GAO-24-106574. In November 2024, OMB created an interagency Territorial Data Gaps Committee to assess the feasibility of including territories in existing statistical products. To fully implement the recommendation, OMB will need to demonstrate how federal statistical agencies, in consultation with territories, are taking steps to better understand territorial data needs and gaps; examining the costs, benefits, and feasibility of including territories in statistical products; and identifying ways to address any data gaps, as appropriate.

needs may be unmet for a variety of reasons. For example, a participant and another agency official we interviewed discussed how the COVID-19 pandemic shutdowns disrupted data production and delayed the release of widely used data, such as the 2020 Census.[61] Delays could also be due to less severe issues, such as waiting to receive source data from territories to produce their Gross Domestic Product estimates.[62]

One participant noted that the unanticipated population growth due to increased immigration in 2023 created a gap between published employment statistics and actual labor market conditions, highlighting the need for more dynamic and responsive data collection mechanisms. Specifically, the Current Population Survey, a dataset for employment statistics, relies on population controls—estimates of total population figures that are updated annually based on Census Bureau projections. These controls are used for weighting survey results, but they may not account for unexpected population changes. The participant cited research[63] that suggests the survey may have underestimated the total

[61]Our work on the 2020 Decennial Census demonstrated how the COVID-19 pandemic shutdowns significantly disrupted the Census Bureau's data collection and quality assurance activities. Operational delays led to the delayed release of the apportionment numbers and redistricting data products. In response to our recommendation, the Census Bureau has taken steps to improve planning for key production, development, and testing schedules. See GAO, *2020 Census: Bureau Released Apportionment and Redistricting Data, but Needs to Finalize Plans for Future Data Products*, GAO-22-105324 (Washington, D.C.: March 2022).

[62]For example, we reported that the availability of Gross Domestic Product estimates for the territories is lagged relative to the rest of the U.S. The data releases were deferred until enough data are available from either territorial governments or other federal statistical products to permit estimation of the Gross Domestic Product. See GAO-24-106574.

[63]Wendy Edelberg and Tara Watson, "The Strong Labor Market Has Boosted U.S.-Born Employment," Brookings Institution (Aug. 13, 2024). https://www.brookings.edu/articles/strong-labor-market-boosted-us-born-employment/.

U.S. population, leading to underestimated employment levels for both U.S.- and foreign-born workers.[64]

Gaps in relevance. One participant from a statistical agency said that policymakers need relevant data and that those needs can evolve, particularly to be responsive and adaptable to rapid change. [65]

One Participant's View on Relevance

"Statistical policy directives… guide us as statistical agencies on what to measure and what's important for policymakers and public and private data users. But as the world changed dramatically and rapidly with the onset of the pandemic, so has the need for a second look at what is important to measure... Users are expecting information on [rapid] change, and they want this information fast and quick…"

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

For example, a participant noted that during the COVID-19 pandemic shutdowns, demand increased for relevant data on the method and mode of learning (e.g., in-person, virtual, or hybrid) for students in primary and secondary schools. At the same time, demand continued for existing data products, such as graduation rates. As such, the COVID-19 pandemic illustrated a broader challenge for statistical agencies to maintain data relevance by balancing the production of existing information with the emerging statistics data users need.

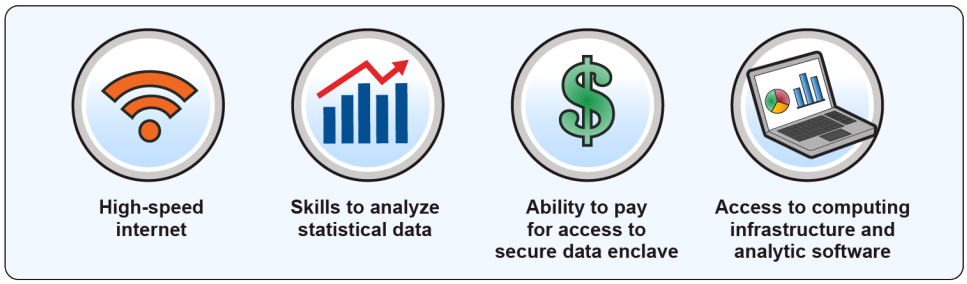
[64]Each year, the Bureau of Labor Statistics updates population controls in the Current Population Survey to align with the Census Bureau's most recent estimates. The January 2025 adjustment, revisions based on 2024 Census data, incorporated significant upward adjustment to net international migration since the 2020 Census base. As a result, the civilian noninstitutional population was revised upward by 2.9 million, and the labor force by 2.1 million. However, according to the agency, "although the effect on levels was relatively large, the effect on [estimated employment] rates and ratios was small." See U.S. Bureau of Labor Statistics, Adjustments to Household Survey Population Estimates in January 2025 (Washington, D.C.: U.S. Department of Labor, February 2025), https://www.bls.gov/web/empsit/cps-pop-control-adjustments.pdf.

[65]In 2024, we reported that national data on evictions could benefit government and nonprofit programs that seek to assist tenants at risk of or experiencing evictions, according to some policymakers and housing advocates. However, the American Housing Survey, which the U.S. Department of Housing and Urban Development sponsored, and the Census Bureau conducted, was not able to provide an accurate estimate of the national rate of evictions in their 2017 survey because of the survey's small sample size. We also reported that no other national estimates of evictions existed. GAO, *Evictions: National Data Are Limited and Challenging to Collect*, GAO-24-106637 (Washington, D.C.: Feb. 28, 2024).

Meeting Diverse Accessibility Needs

According to participants, there are two broad groups of data users—technical and non-technical—who differ significantly in their access to tools, expertise, and resources needed to use federal statistical data effectively. While federal, state, and local governments, businesses, and universities may have both technical and non-technical data users, technical users typically have access to computing infrastructure and analytic software and skills to analyze statistical data (see fig. 4). Non-technical users may not have these resources or skills to analyze statistical data.

Figure 4: Selected Resources That a Technical User of Federal Statistical Data Is Likely to Have



Sources: GAO analysis of the Federal Statistical System Forum, post-forum interview, and National Center for Health Statistics. | GAO-25-107124

According to participants, the federal statistical system faces challenges in addressing the accessibility needs of different data users (see table 8).

Table 8: Reported Challenges for the Federal Statistical System in Meeting Accessibility Needs of Certain Users

Challenge	Description
Streamlined access to restricted microdata	Microdata used for research and policy evaluation are difficult for users to access.
High-performance computing capacity	Users need a high level of computing capacity to efficiently analyze and link large, complex datasets.
Data integration tools	Users need tools that facilitate the integration of statistical data with federal administrative data.
Technical support from the statistical system	Users need support to find, use, and interpret statistical products.
Outreach from the statistical system	Regular input from the broadest range of data users ensures statistical products are relevant.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Streamlined access to restricted microdata. According to participants, technical users may require streamlined access to microdata for detailed research and policy evaluation.[66] For example, researchers analyzing workforce trends make extensive use of microdata from surveys like the Current Population Survey and American Community Survey to study employment patterns, wages, and demographic breakdowns. The ability to access individual-level responses—stripped of personal identifiers—allows researchers to conduct in-depth research, such as obtaining custom estimates that cannot be obtained from pretabulated tables.

According to participants, while agencies produce public-use microdata, access to restricted-use data is limited. Due to confidentiality requirements, access to restricted-use data often requires approval processes and access only within secure data enclaves—making it difficult for researchers to obtain the granular person- or household-level data needed for advanced analysis. In addition, some users may be unfamiliar with the procedures for requesting access to statistical microdata, or may lack the necessary authorization, credentials, or technical infrastructure required to use secure data systems.

High-performance computing capacity. Participants said that current computing capacity of secure data enclaves does not fully serve the needs of technical users beyond the ability to access restricted microdata to create custom estimates. Technical data users need high-performance computing capacity to efficiently access and analyze large, complex datasets, enabling activities such as advanced modeling and linking different datasets. One participant noted that secure enclaves run by the Census Bureau have not fully addressed the capacity needs of technical users—including both the computational resources and system flexibility required for linking large, complex federal and state datasets.

Data integration tools. Participants called for better tools from the federal statistical system that could facilitate integration of state and local administrative records with federal administrative data to inform decision-making—a common need among technical users. According to one participant, states experienced challenges implementing the COVID-19

[66]Microdata are datasets containing information from individual records of people or businesses. To protect the identity of data providers, identifying information is removed from “public-use” microdata. Access may be granted to analyze “restricted-use” microdata—which is not anonymized—however this must be done in restricted, secure settings such as Federal Statistical Research Data Centers.

Pandemic Unemployment Assistance program.[67] Among other things, this program provided unemployment benefit assistance to specified workers who were not covered under existing state or federal unemployment laws, such as self-employed and gig workers. States would have benefitted from federal administrative data—such as historical employment and earnings data for individual applicants—to support unemployment claims but could not easily access these data from the sources accessible to state governments.

In addition, a participant we interviewed told us that some state programs would like to integrate individual-record level data within the state, across states, and between states and the federal government but lack the ability to do so. For example, there is no way to analyze longitudinal changes in employment when people move to a different state because states cannot access the individual-record level data of another state.

Technical support from the statistical system. According to participants, non-technical users may struggle with using the statistical products they need. For example, they may struggle to properly use geospatial data to generate estimates, which can be complex and require specific training, as well as the funds to purchase access to specialized software. Non-technical users may require enhanced support and resources to effectively navigate statistical data products, ensuring they can interpret and utilize available microdata and statistical estimates accurately; however participants told us that providing such support may be costly.

One forum participant noted that localities would like to rely on publicly available estimates from federal survey data for local management tasks, conducting needs assessments, and applying for federal grants. However, non-technical users may encounter obstacles, such as lack of access to necessary software, limited expertise, or lack of tools, such as data visualization dashboards, infographics, and interactive tools, to access these estimates.

[67]See CARES Act, Pub. L. No. 116-136, § 2102, 134 Stat. 281 313 (2020), codified as amended at 15 U.S.C. § 9021.

One Participant's View on Providing Technical Support

"Right now, our statistical agencies primarily disseminate information in two distinct modes, public use statistics and confidential data at research data centers. There is a huge space in between these extremes. How do we develop more intermediate tiers of access that enable more people to have efficient, responsible, and secure access to versions of our data (whether aggregate statistics or extracts of microdata) in new and different ways?"

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

Outreach from the statistical system. OMB policy calls for statistical agencies to seek input regularly from the broadest range of data users.[68] While some statistical agencies regularly engage with technical users—such as researchers—through research conferences, advisory committees, and secure data enclaves, participants told us they may need different strategies to engage with other types of users who may not regularly interact with the system, such as non-technical users.

One Participant's View on Outreach

"We invest a ton of money in data collection and acquisition and in statistical product development. But do we invest enough, and in the right type of resources, in helping stakeholders and users of all sorts (even those we don't know about) who are trying and struggling to use our statistical products?"

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

One participant told us that statistical agencies struggle to connect and assess the needs of non-technical data users, in part because these users may lack the time, resources, or awareness to participate in outreach efforts by statistical agencies. As a result, public engagement may not fully reflect the needs of all users, since certain groups—like researchers and other frequent data users—may have more influence in shaping priorities and in how information is communicated. In addition, limited visibility into how statistical products are used restricts statistical agencies' ability to effectively identify and address the diverse needs of data users.

[68]Office of Management and Budget, Statistical Policy Directive No. 1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units, 79 Fed. Reg. 71610 (December 2014).

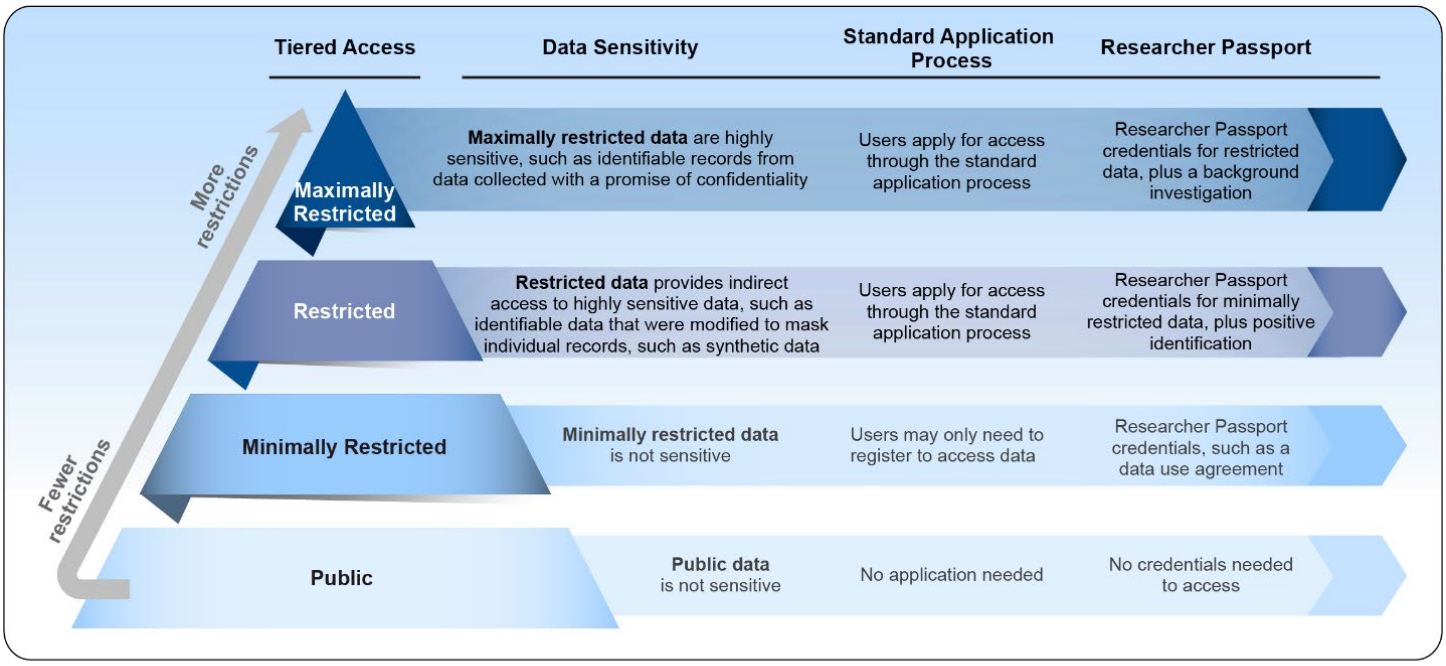
Opportunities to Expand Data Accessibility

User Access

Participants said that a tiered-access model, as envisioned through a potential National Secure Data Service (NSDS), could enhance statistical data access, and could offer varying levels of data access based on users' needs and the sensitivity of the data. The NSDS Demonstration Project, established under the CHIPS Act of 2022, provides a phased approach to test the tools, services, and processes needed to implement such a system.[69] The project is being led by the National Center for Science and Engineering Statistics within the National Science Foundation. Looking beyond the Demonstration Project, participants discussed potential components of a full-scale NSDS, acknowledging that these remain tentative at this stage (fig. 5).

[69]See CHIPS Act of 2022, Pub. L. No. 117-167, § 10375, 136 Stat. 1366, 1574 (2022). Additionally, the Evidence Act included a requirement to establish an Advisory Committee on Data for Evidence Building to review, analyze, and make recommendations to promote the use of federal data for evidence building. See Pub. L. 115–435, § 315, 132 Stat. 5529, 5531 (2019). This committee described the vision and framework for a National Secure Data Service.

Figure 5: Illustration of a Potential National Secure Data Service Tiered Access Framework



Sources: GAO analysis of the Federal Statistical System Forum, Commission on Evidence-Based Policymaking, National Center for Science and Engineering Statistics, OMB Memorandum M-23-04. | GAO-25-107124

Participants said that tiered access could allow agencies to offer multiple levels of access depending on the sensitivity of the data, which would potentially offer more flexibility than the two access levels (public-use and restricted-use) currently available. For example, participants discussed that restricted tiers could include the ability to link statistical data with administrative data and analyze restricted-use microdata at more restricted tiers and access aggregated datasets and use synthetic data for exploratory analysis at less restricted tiers.

In addition to tiered access, a potential NSDS could include a modified approach to the current Standard Application Process[70]—to standardize the process for users to apply for access to data—and Researcher

[70]While a Standard Application Process currently exists, several agency-specific features remain. The process currently has agency-specific eligibility requirements, access modalities and available software, required documents, and project duration limits. The Standard Application Process is available at https://www.researchdatagov.org/.

Passport—to standardize credentials for specific levels of access.[71] Once credentialed, data users with approval to access restricted data from multiple agencies could do so without having to repeat the application and credentialing process for each agency.

For non-technical users, participants said a potential NSDS could offer concierge services to provide a pathway to ask questions about finding, using, and interpreting public federal statistical products through an online portal. Specifically, the services could provide tailored guidance and technical support to help non-technical users navigate complex statistical products based on their specific needs.

Participants noted that as a demonstration project, the NSDS remains untested at scale and successfully implementing it will require sustained resources and funding to ensure its full realization. According to participants, limited funding and staff capacity may hinder the future development of a potential NSDS.

User Outreach and Support

Participants said that, to maximize the utility of statistical products and ensure they are used effectively, agencies should undertake efforts to reach new user groups, provide training, documentation, and guidance for non-technical as well as technical data users, and evaluate existing products (see table 9).

[71]While the Researcher Passport user authentication system, that is the login system, currently exists, the credentialing requirements vary across statistical agencies. Credentialing requirements may include a positive identification of the applicant, confidentiality training, a data use agreement, and a background investigation. See Interagency Council on Statistical Policy, *Standard Application Process Annual Report for 2023* (February 2024), https://ncses.nsf.gov/1059/assets/0/files/sap-annual-report-2023.pdf, accessed April 1, 2025.

Table 9: Reported Opportunities for the Federal Statistical System to Enhance User Outreach and Support

Opportunity	Description
Expand outreach	Expand outreach to new data user groups.
Provide training	Provide training on how to find, use, and interpret federal statistical products.
Provide documentation and guidance	Provide accessible documentation and guidance, such as data dictionaries and user guides, and allow users to access them through various platforms, such as chatbots.
Evaluate existing products	Make strategic decisions about discontinuing certain products and pivoting to new products to meet user needs.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Expand outreach. Participants told us that expanding outreach to data users, identifying new user groups, and recognizing emerging needs is an important part of the mission of the federal statistical system. One participant identified the Trust Regulation, which includes an opportunity to ensure that the federal statistical system is responsive to data users, such as policymakers and researchers.[72] The Trust Regulation requires recognized statistical agencies and units to maintain regular understanding of the needs and interests of data users, and as appropriate, consult with other data users to assess and seek improvements to the relevance of statistical products through user outreach.[73]

One participant provided an example of outreach between statistical agencies and users, where statistical agencies were responsive to user input. The Economic Research Service within the U.S. Department of Agriculture participates with the National Agricultural Statistics Service in a bi-annual Data Users Meeting, which is an open forum for data users to ask questions about the agencies' statistics programs. According to this participant, from a customer service perspective, this meeting provides an opportunity to learn about data users' concerns and desires for improvements or changes to the statistics and economics programs at these agencies.[74] Furthermore, participants said that identifying new

[72]5 C.F.R. § 1321.5(b).

[73]5 C.F.R. § 1321.5(b)(1)(ii).

[74]In addition, this participant noted that as part of its *Democratizing the Data* initiative, the U.S. Department of Agriculture's Economic Research Service hosted a 2023 workshop to gather community input on data usage statistics in support of the Foundations for Evidence-Based Policymaking Act. The event featured hands-on sessions with tools like usage dashboards and encouraged participants to explore how data are used in scientific and public research. See *Harvard Data Science Review*, https://democratizingdata.ai/.

groups of data users, particularly non-technical users, is important to fostering a more comprehensive and representative data ecosystem and increase the value of statistical products.

One Participant's View on Finding New User Groups

"Statistical agencies have to go out and find out what our users really want. And I don't just mean the people that are using data products heavily, I mean new users we want to attract."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

Provide training. Participants said that statistical agencies could provide ongoing trainings designed for non-technical users to use their statistical products. For example, one participant told us that the Economic Research Service at the Department of Agriculture provides a data training series led by the agency's subject matter experts. These trainings cover various topics, including agricultural productivity, commodity-specific data, and economic indicators.[75] While the content is often detailed, the webinars are available to the public and aim to make complex data accessible to a broad audience, including both technical and non-technical users.

In addition, the Census Bureau provides free, public courses through Census Academy, including training for technical users on how to use Application Programming Interfaces—tools that allow different software systems to communicate—to integrate Census data into their own applications.[76] For non-technical users, the Academy also offers short video tutorials on how to explore housing trends using Census data tools. Another participant told us that the Bureau of Economic Analysis at the Department of Commerce regularly conducts webinars for non-technical users and hands-on training at conferences for technical users.

Provide documentation and guidance. According to participants, accessible documentation and guidance designed for non-technical users would help them navigate and use federal data products more

[75]"Past Data Training Webinars," Data Training Webinars, Economic Research Service, accessed February 20, 2025, https://www.ers.usda.gov/newsroom/trending-topics/data-training-webinars.

[76]Census Academy, Census Bureau, last updated April 9, 2025, https://www.census.gov/data/academy.html.

effectively.[77] In a follow-up interview, an agency official said that their office recognizes opportunities to improve the transparency of its data documentation and expressed the importance of "showing our math," that is defining the data used in statistical products clearly and precisely through publicly accessible documentation, such as user manuals and data dictionaries. Other participants said that the statistical system could provide guidance through chatbots, developing crosswalks of data definitions and variables across agency survey and administrative data, and making data documentation easier for users to access and understand.

Evaluate existing products. With constrained resources, participants said statistical agencies increasingly need to make strategic decisions about discontinuing certain products and services and pivoting to new products and services to meet user needs. A participant suggested that agencies could obtain public input on discontinuing products from advisory committee meetings and other public panels, and this outreach could help augment product download statistics that may not provide a complete picture of how users interact with statistical data.[78] In doing so, statistical agencies could determine how their products are used and confirm their value to data users.

When evaluating whether existing statistical products meet user needs, participants encouraged statistical agencies to regularly assess how existing products can be improved or supplemented through the integration of alternative data sources. This includes assessing whether administrative records, commercial data, or sensor-based inputs can enhance the timeliness, granularity, and relevance of statistical outputs to better align with user needs. Based on such assessments, participants told us agencies could choose to prioritize or deprioritize certain products, enable rapid innovation, align resources with data user priorities, and

[77]A report by the National Academies noted that without access to clear documentation of methods and data, the credibility of federal statistics could be compromised, potentially reducing their perceived reliability and utility. See National Academies of Sciences, Engineering, and Medicine, *Transparency in Statistical Information*, 2022.

[78]In 2014, the Census Bureau reviewed federal agency uses of every question on the American Community Survey questionnaire. Based on this review, the Bureau categorized each question by its utility and respondent burden, providing the Bureau with justification to remove questions with low utility from the 2016 American Community Survey. See U.S. Census Bureau, "2014 Content Review" (last updated June 30, 2015), https://www.census.gov/programs-surveys/acs/methodology/archived/2014-content-review.html.

ensure agency efforts contribute meaningfully to informed decision-making.

Alternative Data Could Help Modernize Statistical Production and Meet User Needs, but There Are Challenges to Navigate

Traditional statistical data have limitations, and forum participants highlighted key advantages that alternative data sources offer for federal statistical production and meeting the needs of its users. However, participants said that statistical agencies face significant challenges in accessing and using alternative data, such as limited federal authority to access private-sector data and complications stemming from the non-statistical origins of alternative data. Participants said that providing incentives for administrative data providers and building a robust confidentiality infrastructure to address disclosure concerns represent opportunities for integrating alternative data into statistical production.

Alternative Data Could Help Improve Statistical Production and Meet User Needs

Participants said that alternative data can help meet growing user demand for granular, timely, accurate, and complete statistical products, as well as improve the efficiency of federal data collection efforts (see table 10).[79]

Table 10: Reported Benefits of Using Alternative Data in Federal Statistical Production

Benefits	Description
Granularity	Some alternative data, such as transaction-level or county-level data, may be more granular than survey data.
Timeliness	May be accessed and analyzed more frequently than survey data, and sometimes in real-time.
Augment survey data	May be used to crosscheck survey responses for accuracy and improve data completeness.
Efficiency gains	May be more efficient to obtain than survey data.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

[79]A vision for modernizing the U.S. data infrastructure consists of having the tools to use alternative data to improve the quality, timeliness, granularity, and usefulness of national statistics and support research, policymaking, and program evaluation, according to a National Academies panel. See *Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good*, 2023.

Experimental Data Products

Data linkage, also known as data matching or entity resolution, connects and integrates data from multiple sources by matching record-level information such as individuals or businesses. This technique has been used to advance experimental data products.

For example, the Monthly State Retail Sales is the Census Bureau's new experimental data product featuring modeled state-level retail sales. This is a blended data product using Monthly Retail Trade Survey data, administrative data, and private-sector data. The initiative aims to enhance the timeliness of retail sales data, addressing user demand for more current economic indicators.

Source: Thitichaya/stock.adobe.com (image); GAO analysis of agency documentation. | GAO-25-107124

Granularity. According to participants, granularity can be a key advantage of both administrative and private-sector data. For example, one participant said that administrative data from states, localities, and Tribes can "build the capacity of the country to use and produce high quality data at different geographic levels." Another participant credited private-sector data with allowing the Bureau of Economic Analysis to "pinpoint where oil and gas production is coming from at the county level," part of the agency's methodology for measuring Gross Domestic Product by county. The latter participant identified granularity as the most notable opportunity afforded by alternative data.

Timeliness. Frequency and timeliness—sometimes even real-time data feeds—were distinct advantages that participants attributed to private-sector data. For participants, this advantage was particularly apparent with economic data, such as retail transactions. One participant told us, "Businesses are tracking, in real time, prices and quantity and attributes in ways that current statistics don't take advantage of." Another participant described how the Census Bureau has started supplementing its Economic Census—a 5-year measure of American businesses—with more-frequently updated data from a company that aggregates proprietary data from different sources.

One Participant's View on Timeliness

"…instead of doing a survey, [the Census Bureau is] getting a direct feed. … This is a way to allow us to efficiently get the data, curate the data, make it available to our [Monthly State Retail Sales team] to be able to [produce] the data products they need to do, be able to provide data to … our data users in an efficient, documented, repeatable fashion."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

A third participant explained how access to private-sector credit card data has helped the U.S. Bureau of Economic Analysis include more timely estimates of consumer spending in its Advanced Estimate of the Gross Domestic Product, which is published in the first month after each quarter, and more timely consumer spending estimates at the subnational level.

Augment survey data. Participants also described how alternative data—especially administrative records from high-coverage databases—may be used to crosscheck survey responses for accuracy, fill in gaps caused by survey non-response, and measure important attributes that were outside of a survey's scope. For example, participants from the research community noted that some sources of income, such as

pensions and food assistance benefits, are often misreported, leading to inaccuracies in economic and social policy analyses. The participants discussed research projects which aim to supplement survey data by linking with federal administrative records to improve the completeness and granularity of income and poverty data.[80] Participants also saw potential in private-sector data to improve data completeness and accuracy.

Efficiency gains. According to participants, administrative data have the potential to make data collection more efficient and cost-effective by reducing statistical agencies' reliance on costly forms of direct data collection. Participants said that statistical agencies might be able to shift some of the data collection burden away from costly surveys and censuses by using datasets that already exist.

For example, one participant shared how the U.S. Department of Agriculture's Economic Research Service improved efficiency by integrating the U.S. Geological Survey's Protected Area Database with the statistical agency's Major Land Uses product—a census that classifies all acres in the United States. The participant explained that it is particularly difficult to collect data on "special use areas," such as rural parks, wilderness areas, wildlife areas, and defense and industrial areas, due to the complexity of land ownership. For the Economic Research Service, the involvement of state, local, and private landowners has traditionally required "an extremely broad and labor-intensive" data collection effort across state and local data sources. However, by accessing geospatial data in the Protected Area Database, the Economic Research Service has been able to abandon its expensive traditional data collection effort while still obtaining the same information, thereby saving resources.

[80]See, for example, the Comprehensive Income Dataset—a University of Chicago initiative to link together federal survey and tax data with administrative records from federal and state governments. The Census Bureau has a similar initiative called National Experimental Well-Being Statistics, which links together survey, decennial census, administrative, and private-sector data.

Agencies Face Challenges Related to Privacy, Data Sharing, and Reliability when Integrating Alternative Data

Participants identified various challenges that make it difficult to integrate alternative data into federal statistical production (see table 11).

Table 11: Reported Challenges to Using Alternative Data in Federal Statistical Production

Challenges	Description
Data provider concerns about data protection	Administrative data may lack strong confidentiality protections as well as clear disclosure statements to data providers about how their data might be used for statistical purposes.
Barriers to using private-sector data	Agencies may have difficulty obtaining and using privately owned data.
Legal constraints on administrative data sharing	Legal barriers may prevent, narrow, or delay access to government administrative data.
Reliability of alternative data for statistical purposes	The non-statistical origins of alternative data may bring data reliability risks.
Human capital demands	Alternative data may require expertise and skills that are less common in statistical agencies.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Data provider concerns about data protection. Unlike federal statistical data, administrative data provided to the federal government may lack strong confidentiality protections as well as clear disclosure statements to data providers about how their data might be used for statistical purposes. Participants said this may make some potential data provider groups—such as states, localities, and federally recognized Tribes—or government program participants hesitant to share their data with federal agencies as a whole.[81] According to one participant, hesitancy may be fueled in part by a lack of trust in federal agencies to safeguard the raw data.

Concerns about the current legal framework for data protection may also play a role. According to one agency official we interviewed, Tribes have been reluctant to share administrative data, such as school enrollments and student microdata, with the Bureau of Indian Education, which

[81]Our work in the area of cybersecurity highlighted that the federal government could do more to improve the protection of federally collected and maintained personal and sensitive data. See *Cybersecurity High-Risk Series: Challenges in Protecting Privacy and Sensitive Data*, GAO-23-106443 (Washington, D.C.: Feb. 14, 2023).

Digital Agriculture and Private-Sector Data

According to one participant, digital agriculture, which uses automation and data to transform farming, brings both opportunities and challenges for integrating private-sector data into the federal data infrastructure. This participant said that tensions may emerge when firms collect on-farm data for research while farmers seek to maintain agency over their data. The National Cooperative Dairy Herd Improvement Program, a partnership among dairy farms, genetics companies, and the U.S. Department of Agriculture, illustrates this dynamic. The program collects animal-level data to improve bull genetics evaluations and facilitate the flow of information in the market. According to this participant, the program's goals raise critical questions regarding what services can incentivize farmers to share data, how data standards can be coordinated across public and private platforms, and how farmers' data rights can be protected while supporting innovation.

Source: Clara/stock.adobe.com (image); GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

sometimes shares data with statistical agencies.[82] The official said that Tribes are sensitive about potential public disclosure of data through records requests under the Freedom of Information Act.[83] This official said that the concern is rooted in Tribes' loss of control over the data once they are in the hands of federal agencies.

Barriers to using private sector data. Participants also highlighted that agencies may have difficulty obtaining and using private-sector data. For example, technology companies capture large volumes of data on the activities of children and teachers in public classrooms. However, these data are owned by the companies, not by the public—which may limit the access of federal statistical agencies like the National Center for Education Statistics. In addition, one participant explained how agencies sometimes find it difficult to extract greater value from the private-sector data they procure. The participant provided an example of partnerships between statistical agencies and university researchers, in which researchers' institutions may be reluctant to sign data use agreements required by the private-sector data provider.[84] These comments suggest that it is often hard to align federal priorities with those of private-sector data providers, which can make collaborations to share alternative data sources less effective.

[82]For example, the Common Core of Data—the primary database on public elementary and secondary education managed by the National Center for Education Statistics—includes data submitted annually by schools administered by the Bureau of Indian Education. See "Common Core of Data (CCD)," NCES Handbook of Survey Methods, National Center for Education Statistics, accessed on June 30, 2025, https://nces.ed.gov/statprog/handbook/ccd.asp.

[83]The Freedom of Information Act established a statutory right of public access to certain federal agency records, subject to exemptions, including personal privacy, national security, and law enforcement interests. Pub. L. No. 89-487, 80 Stat. 250 (1966), codified at 5 U.S.C. § 552. See also, definition of agency 5 U.S.C. § 551.

[84]When statistical agencies work with university partners, they may need special agreements to share private-sector data. According to one participant, the compliance requirements in these agreements can be burdensome for universities, sometimes making them reluctant to collaborate.

Participants said that agencies' dependence on negotiating agreements with the private sector to access their data may introduce uncertainty in the cost and consistent availability of agencies' data sources.[85] Participants expressed concerns that agencies would come to rely on particular private-sector data sources, which could incentivize these private-sector data providers to leverage this dependence and increase their prices. Moreover, private-sector data sources sometimes disappear for reasons not always clear to statistical agencies—for example, due to a provider going out of business or an opaque refusal by a provider to continue selling or sharing its data.

Legal constraints on administrative data sharing. Participants said legal authorities for access to government administrative data can be unclear, or can prevent, narrow, or delay statistical agencies' access to these data. For example, participants cited explicit legal restrictions on agencies' access to confidential tax data[86] from the Internal Revenue Service for statistical use.[87] In other cases, conflicting interpretations of applicable statutes can contribute to confusion and delay. One participant shared an example of an attempt by the Census Bureau to negotiate state-by-state data sharing agreements for Supplemental Nutrition Assistance Program data, regulated by the U.S. Department of Agriculture. According to this participant, some states believed that existing an statute allowed them to share these data for statistical purposes, while other states read the same statute in the opposite way. The data sharing agreements were able to proceed only after the U.S. Department of Agriculture confirmed that these agreements were legal.

[85]According to one participant, private-sector data acquisition costs increase over time, and statistical agencies like the Economic Research Service at the U.S. Department of Agriculture rely on private-sector data for research purposes. While there are a set of private-sector data products that provide specific data consistently over time, the greater challenge for the Economic Research Service is the consolidation of proprietary data over time. Specifically, the same firm may bundle sought-after databases with other data products, thus increasing annual acquisition costs for statistical agencies.

[86]That is, confidential tax returns and return information.

[87]See 26 U.S.C. §6103(j). Neither the U.S. Bureau of Labor Statistics nor its parent agency, the U.S. Department of Labor, are provided with the legal authority to use disaggregated Internal Revenue Service tax data for statistical purposes. Meanwhile, the U.S. Bureau of Economic Analysis has authority to access tax return information of corporations only to the extent necessary for structuring censuses and national economic accounts and conducting related statistical activities authorized by law. It has no authority to access return information of noncorporate businesses. Noncorporate businesses include sole proprietorships and partnerships.

A participant also cited the Census Bureau's Longitudinal Employer-Household Dynamics program, which collects Unemployment Insurance earnings data and Quarterly Census of Employment and Wages data from states and territories to enhance labor market statistics. The participant noted that national coverage requires separate legal negotiations and agreements with individual states and territories, a process that is still ongoing after more than 25 years.[88]

Reliability of alternative data for statistical purposes. Participants highlighted how blending data from survey and non-survey sources can raise issues with data reliability. Survey data are designed and collected to obtain an unbiased estimate with a desired level of precision. In contrast, alternative sources, such as administrative data, are originally collected for non-statistical purposes, such as ensuring compliance with program requirements, and not for comparability across time or calculating precise estimates. Because of these origins, participants said that incorporating alternative data may pose complications, the foremost being impaired or unknown data reliability. According to one participant, this can happen due to lack of oversight over key methodological decisions.

One Participant's View on the Reliability of Alternative Data Sources

"There's little insight into, or control over, how [alternative] data are collected. The access and mode of collection change over time, in ways that can undermine the longitudinal consistency of statistical data."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

While some alternative data sources can be used to provide more complete and timely statistics, other alternative data sources can pose challenges. Participants said that administrative datasets are not immune to incomplete or inaccurate information that may be submitted by data providers. In addition, one participant highlighted how inconsistent reporting requirements across various federal agencies may overextend state and local data administrators and affect the general quality of data they provide, especially if they struggle to see benefits beyond fulfilling reporting requirements.

[88]The Longitudinal Employer-Household Dynamics program was launched in 1999. Currently, 46 states, one territory, and the District of Columbia participate. See "State Partners," Longitudinal Employer-Household Dynamics, U.S. Census Bureau, last updated August 5, 2025, https://lehd.ces.census.gov/state_partners/.

Human capital demands. According to participants, another complication of data with non-statistical origins is the unique resource demands. In particular, statistical agencies must invest significant effort to process data for uses beyond its original intended purpose. Participants also pointed out that processing and analyzing alternative data may require expertise and skills (e.g., in data engineering and data science) that are less common in statistical agencies where survey methodology has been the norm.

In data sharing agreements, agency attorneys are responsible for developing terms and conditions to document the agreement, a process that often involves complex, time-intensive negotiations and careful compliance with statutory and regulatory requirements. According to participants, the limited capacity of agency attorneys can contribute to bottlenecks, which can be a barrier to acquiring and using alternative data.

Better Incentives and Confidentiality for Administrative Data Providers Could Help Integrate Administrative Data Sources

Participants said that creating incentives for administrative data providers and enhancing confidentiality protections are key ways that the federal government could modernize statistical production by making it easier to integrate data from different sources into the system (see table 12).

Table 12: Opportunities to Integrate Administrative Data in Federal Statistical Production

Opportunities	Description
Incentivize administrative data providers	Providing incentives, such as useful analyses of data, could encourage state and local governments to share administrative data.
Mitigate disclosure risk and ensure transparency	Statistical agencies could improve the availability and reliability of administrative data sources by mitigating disclosure risk and being more transparent about how data will be used.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Enhancing State and Local Capacity

According to participants, statistical agencies can further incentivize states and localities to share data by helping them build their own technical capacity and data infrastructure. For example, the State Justice Statistics Program, operated by the U.S. Bureau of Justice Statistics, provides limited funds for states to coordinate their statistical activities and conduct research, to maintain and enhance each state's capacity to address criminal justice issues. In exchange, states serve a liaison role in assisting the Bureau to gather data from respondent agencies within their states.

Source: Mandritoiu/stock.adobe.com (image); Bureau of Justice Statistics. "State Justice Statistics Program," Programs & Initiatives, accessed May 30, 2025, https://bjs.ojp.gov/programs/state-justice-statistics-program and GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Incentivize administrative data providers. Participants recommended providing incentives for state and local governments to share administrative data with federal agencies. Receiving useful statistical products from statistical agencies after providing data could serve as a meaningful incentive for providers to share their data. A participant told us that state and local governments are "ground zero" for providing alternative data, such as administrative data.

For example, according to another participant, the Census Bureau's Supplemental Nutrition Assistance Program Eligibility and Access interactive data visualization links the Census Bureau's American Community Survey data with state administrative records to model estimates of food assistance benefits eligibility and access rates at the state and county levels.[89] The participant said that providing information about these programs encourages states to share their administrative data, because they benefit from knowing who receives benefits. This coordination may increase meaningful exchanges, where data providers provide their time and responses in return for data produced by statistical agencies. As an example of these mutually beneficial exchanges, one participant described their agency's efforts to obtain real-time data on COVID-19-related deaths in states and territories during the pandemic:

One Participant's View on Incentivizing Data Providers

"These efforts turned out to be remarkably successful. And I think that's largely because the states and territories saw the value in providing their data, not only to build … national-level aggregates, but also...they saw them as a tool for their own state and local policymaking needs. Data came to us, we added value to it, we shared it back with them, they used it for their purposes."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

Mitigate disclosure risk and ensure transparency. According to participants, statistical agencies could improve the availability and reliability of administrative data sources by addressing sources of distrust among government agencies as well as government program participants who contribute data to administrative systems. One source of distrust is

[89]See Supplemental Nutrition Assistance Program (SNAP) Eligibility and Access (visualization), U.S. Census Bureau, last updated March 4, 2025, https://www.census.gov/library/visualizations/interactive/snap-eligibility-access.html.

the risk of disclosure of private information.[90] Some participants said that proactively building a robust infrastructure to share data in secure environments could be one way to alleviate privacy concerns, foster trust, and incentivize information sharing.[91] Such infrastructure could include supporting a potential National Secure Data Service (NSDS) and related initiatives.[92]

Other participants highlighted the importance of transparency and its role in establishing the consent of parties involved. One participant noted that acquiring a non-statistical agency's consent to use its administrative data for particular statistical uses—such as sharing unprotected, identifiable data with researchers—could prevent breaches of trust and maintain that agency's future cooperation in providing administrative data to the partnering statistical agency. Another participant stated that federal survey data collection standards could serve as a model for establishing the consent of government program participants to use their administrative data for statistical purposes. When collecting data directly

[90]The confidentiality of data has a direct effect on the integrity of the data. Integrity refers to "the maintenance of rigorous scientific standards and the protection of information from manipulation or influence as well as unauthorized access or revision." See Federal Committee on Statistical Methodology, *A Framework for Data Quality*. (Federal Committee on Statistical Methodology, 2020), 3.

[91]In 2017, a panel convened by the National Academies of Sciences, Engineering, and Medicine issued two reports discussing ways to preserve privacy and promote data sharing. The reports note that data providers trust statistical agencies to protect their data, but also that data sharing increases the risk of privacy breaches. To mitigate this risk, they recommend that statistical agencies rely on computer science technologies, such as modern database and cryptography technologies; statistical methods, such as adding noise to individuals' responses; and administrative procedures, such as rules for restricting access. In addition, technical staff of statistical agencies should receive appropriate training in these technologies. See, National Academies of Sciences, Engineering, and Medicine. *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps.* (Washington, D.C.: The National Academies Press, 2017), doi.org/10.17226/24893 and National Academies of Sciences, Engineering, and Medicine. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. (Washington, D.C.: The National Academies Press, 2017), doi.org/10.17226/24652.

[92]According to one participant, one way to minimize risks would be to pseudonymize all data that a statistical agency collects or receives. This participant suggested that only a very limited number of statistical officers should see non-pseudonymized data or know keys for pseudonymization. Similarly, another participant cited the example of using protected identification keys to facilitate data linkage without using sensitive information like name, social security number, or date of birth. Protected identification keys are unique identifiers probabilistically assigned to administrative and survey data held by the Census Bureau to allow for secure linkage of datasets. See "Data Ingest and Linkage," Technical Documentation, U.S. Census Bureau, last modified December 16, 2021, https://www.census.gov/about/adrm/linkage/technical-documentation/processing-de-identification.html.

from individuals—such as through surveys conducted by statistical agencies—federal agencies are required to clearly communicate the purpose of the data collection, which may include programmatic or research goals.[93] In addition, agencies are required to provide a clear explanation of the authority for the collection, the principal purposes for which the data will be used, routine uses, and the effects, if any, of not providing the data.[94] These laws prioritize transparency at the point of data collection, but do not typically establish mechanisms for securing informed consent for secondary statistical use of administrative records.

Improved Interagency Coordination Could Make the Statistical System More Effective and Efficient, but There Are Challenges to Navigate

Participants said that effective interagency coordination is key for modernizing statistical production, facilitating outreach to users, and alleviating resource constraints.[95] According to participants, the decentralized design of the system and the absence of shared regulatory frameworks hinder coordination among agencies, creating barriers to data sharing. Participants said that strengthening interagency coordination requires leveraging existing frameworks and implementing new reforms.

Improved Interagency Coordination Could Make the Statistical System More Effective and Efficient

Participants identified several ways that improved interagency coordination could help the statistical system better meet the needs of data users while alleviating resource constraints exacerbated by the decentralized design of the system (see table 13).

[93]See 5 U.S.C. § 552a(e)(3).

[94]See 44 U.S.C. § 3506(c)(1)(B)(iii).

[95]For example, the National Center for Science and Engineering Statistics coordinates with other statistical agencies to understand and use existing data for cross-agency purposes, coordinate and plan new data collections, and standardize data collection measures. In addition, identifying lead agencies or individuals and clarifying their roles and responsibilities can foster strong and sustained leadership, helping maintain interagency efforts over time. See GAO, *Science and Engineering Statistics: Improved Communication Needed with Stakeholders on Data Needs*, GAO-23-106361 (Washington, D.C.: Sept. 27, 2023) and GAO, *Government Performance Management: Leading Practices to Enhance Interagency Collaboration and Address Crosscutting Challenges*, GAO-23-105520 (Washington, D.C.: May 24, 2023).

Table 13: Reported Benefits to Improving Coordination Among Federal Statistical Agencies

Benefit	Description
Enhanced human capital	Making job positions more transferable between statistical agencies and cultivating professional relationships across statistical agencies may help recruit talent and build collective expertise.
Streamlined acquisition of private-sector data	Coordinating the purchasing of datasets may avert multiple agencies from buying the same or similar datasets, reducing costs.
Interagency trust in data sharing	Communicating clearly and consistently may foster trust among federal agencies, affecting willingness to collaborate in data sharing.
Innovative products	Collaborating across statistical agencies to provide innovative statistical products may help meet users' data needs.
Seamless and integrated data dissemination	Shared services may promote a more seamless and integrated experience for data users, including statistical agencies.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Enhanced human capital. Participants saw staffing and expertise as important for modernizing the statistical system, but the current decentralized framework often results in fragmentation of specialized knowledge and inefficiencies in workforce development. For example, one forum participant noted that junior talent across the government is siloed, and better coordination could help build social capital and improve interagency trust. Another participant referred to the recent Office of Personnel Management guidance on cross-government job application sharing, which aims to improve the quality of applicant pools.[96]

Streamlined acquisition of private-sector data. Participants reported that multiple agencies independently procure the same or similar datasets from private-sector vendors—such as curated market research data and financial information—possibly leading to increased costs. Participants said interagency coordination could make the federal statistical system more efficient by identifying redundant purchases, enabling interagency sharing of purchased data, and improving federal purchasing power. However, one participant cautioned against making procurement "overly centralized," as individual statistical agencies are sometimes able to strike innovative agreements with private-sector companies.

Interagency trust in data sharing. Participants saw coordination as a way to foster trust and improve data sharing among agencies—for

[96]Office of Personnel Management, *Guidance to Support Cross-Government Applicant Sharing,* https://www.opm.gov/media/miqnneng/guidance-to-support-cross-government-applicant-sharing.pdf.

example, through consistent and transparent communication. According to participants, a key historical example of trust erosion and recovery within the federal statistical system occurred between the Internal Revenue Service and the Census Bureau. Consistent with statutory requirements, the Census Bureau was granted access to confidential tax data from the Internal Revenue Service for statistical purposes. However, during a regular review by the Internal Revenue Service in 1999, it was found that the Census Bureau's uses of tax data extended beyond what was previously discussed and documented between the two agencies. Participants said that this incident damaged interagency trust and made the Internal Revenue Service hesitant to expand data sharing with the Census Bureau. To restore the relationship, both agencies implemented stronger governance and oversight mechanisms, including stricter data-use agreements and improving interagency communication.

Innovative products. Participants saw interagency coordination as a key enabler for innovative statistical products that meet evolving data needs. For example, the U.S Census Bureau's Household Pulse Survey was a collaborative interagency effort that demonstrated the federal statistical system's ability to rapidly pivot and coordinate across agencies in response to the COVID-19 pandemic shutdowns.[97] According to participants, the survey's experimental and adaptive design set a new precedent for agile, cross-agency data collection and sharing in times of national emergency.

Seamless and integrated data dissemination. Participants saw interagency coordination as essential to building a more seamless and integrated experience for data users. Shared service initiatives—like Federal Research Data Centers, the Standard Application Process, and the potential NSDS—can enhance data linking capabilities, promote use of privacy preserving technologies, and improve data accessibility for all users, both inside and outside the federal government. This includes statistical agencies, which may exchange data with each other or integrate data from other federal sources to improve statistical products.

[97]According to the Census Bureau, the Household Pulse Survey is designed to provide relevant data with a short turnaround time; however, as part of the agency's Experimental Data Series, this product may not meet some of the agency's statistical quality standards. In January 2025, the Household Pulse Survey was relaunched as the Household Trends and Outlook Pulse Survey (HTOPS). See "Household Pulse Survey: Measuring Emergent Social and Economic Matters Facing U.S. Households," Experimental Data Products, U.S. Census Bureau, last modified April 30, 2025, https://www.census.gov/data/experimental-data-products/household-pulse-survey.html.

Decentralization, Single Use Data Sharing, and Data Privacy Considerations Hinder Effective Interagency Coordination

Forum participants said that many statistical agencies are limited in their abilities to coordinate efforts among themselves and with other federal agencies to pursue system-wide goals. Participants identified several challenges that hinder effective interagency coordination, particularly as it relates to streamlining data procurement and interagency data sharing (see table 14).[98]

Table 14: Reported Challenges Hindering Effective Coordination Among Federal Statistical Agencies

Challenge	Description
The system's decentralized design	Statistical agencies are housed within different parent agencies, creating inefficiencies for coordination.
Time-consuming interagency data use agreements	The federal government lacks a shared framework for interagency data sharing, and data sharing agreements are often on a project-by-project basis. This may lead to delays and administrative burden.
Confidentiality-related constraints on interagency data sharing	Certain considerations, including statutory protections for data confidentiality, sometimes restrict agencies' ability to share data with each other.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

The system's decentralized design. Forum participants noted that the decentralized design of the federal statistical system allows statistical agencies to specialize in specific policy areas of its parent organization, but it also creates inefficiencies for effective collaboration among statistical agencies and between statistical agencies and other federal agencies. According to one participant, the federal statistical system was not originally structured around shared priorities, regulatory frameworks, and services. Instead, the system was built upon distinct operational

[98]We have previously written about the importance of administrative data sharing for maximizing the efficiency of existing data sources. For example, in 2012, we recommended that the Office of Management and Budget (OMB) develop comprehensive guidance for both statistical agencies and agencies that hold administrative data, to use when evaluating and negotiating data sharing. In 2014, OMB implemented this recommendation by issuing Memorandum M-14-06, which provides agencies with guidance for addressing the legal, policy, and operational issues that exist with respect to using administrative data for statistical purposes. However, additional action can sometimes be taken at the agency level, as shown by our open 2022 recommendation that the Department of the Treasury do more to evaluate the feasibility of data sharing. See "Federal Statistical System: Agencies Can Make Greater Use of Existing Data, but Continued Progress Is Needed on Access and Quality Issues," Reports & Testimonies, U.S. Government Accountability Office, accessed July 21, 2025, https://www.gao.gov/products/gao-12-54, and "Tax Equity: Lack of Data Limits Ability to Analyze Effects of Tax Policies on Households by Demographic Characteristics," Reports & Testimonies, U.S. Government Accountability Office, accessed July 21, 2025, https://www.gao.gov/products/gao-22-104553.

structures, distinct regulatory approaches, and service-specific frameworks.[99]

Participants said existing laws enable coordination within this decentralized system but may not address all challenges. For example, although the Paperwork Reduction Act of 1995 streamlined the federal government's collection of information from the public,[100] one participant noted that the law does not apply to private-sector datasets.[101] The absence of similar oversight mechanisms for private data procurement makes it difficult to understand the prevalence of duplicative procurement and how it might be streamlined.

Time-consuming interagency data sharing agreements. According to participants, another reason why it is difficult for federal agencies to share statistical or administrative data with one another is that the process of negotiating a data sharing agreement can be costly in terms of effort and time. When contacted after the forum, one participant noted that the federal government lacks a shared framework for data sharing, leading to duplicative and time-consuming negotiations between agencies as they develop data use agreements from the ground up. These costs accumulate when data sharing agreements apply only to a particular use, leading to repeated negotiations between agencies. Participants said that single use interagency data sharing agreements may delay progress and lead to administrative burden.[102]

[99]For an overview of the history and structure of the Federal Statistical System, see National Academies of Science, Engineering, and Medicine. *Principles and Practices for a Federal Statistical Agency: Eighth Edition*, Appendix B.

[100]Under the Paperwork Reduction Act of 1995, OMB is responsible for reviewing and approving federal information collection requests to avoid unnecessary duplication, reduce respondent burden, and ensure the practical utility of collected data. However, the scope of this oversight is limited to collections of information from the public and does not extend to the procurement of data from commercial vendors. See Pub. L. No. 104–13, 109 Stat. 163 (1995).

[101]Pub. L. No. 104–13, 109 Stat. 163 (1995).

[102]In 2023, we reported that data collection on nonstandard work arrangements is fragmented across seven federal agencies, including statistical agencies. We found that this fragmentation contributes to limitations in data quality and results in inconsistent estimates that are not directly comparable to each other. One reason for this is agencies' priority on the needs of individual programs rather than crosscutting efforts across the system. We recommended that there should be an interagency collaborative mechanism to improve these measurements. See GAO-24-105651.

In post-forum comments, participants explained that privacy protection is a prominent motivation for this; even projects using the same dataset are evaluated separately for appropriate statistical purposes, safeguards, and access rights consistent with data provider consent. Other participants said that data use agreements require explicit specification of the data's intended uses to ensure compliance with applicable statutes that contain use limitations, thus precluding agreements with broad applications.[103] One participant also suggested that agencies' tendencies to adhere to established norms could hinder them from adopting new data sharing practices.

Confidentiality-related constraints on interagency data sharing. Participants stated that interagency sharing of statistical and administrative data is often in tension with the priority of confidentiality protection, with the latter sometimes restricting the former through statutory provisions. Some statistical agencies, such as the Census Bureau, have limited access to tax records from the Internal Revenue Service under a specific statutory provision that regulates the confidentiality and disclosure of tax data, but other statistical agencies do not have access to these data at all.[104] Another statute restricts the Census Bureau from sharing individual reports.[105] Such sharing may result in the identification of an individual or business.

Participants noted that these restrictions can hinder statistical agencies that might otherwise use Internal Revenue Service or Census data to create sampling frames or make their own statistics more consistent and comprehensive. For example, the Bureau of Labor Statistics cannot access business data reported on individual tax returns, which limits its ability to efficiently and reliably produce labor statistics. At the same time, participants also noted the importance of ensuring robust safeguards to protect the confidentiality of these data, as existing laws were established to address risks of misuse, unauthorized access, and breaches of confidentiality.

[103]For example, the Higher Education Act of 1965, as amended, generally restricts the use of information provided by an applicant for the Free Application for Federal Student Aid to the application, award, or administration of a federal aid program. Pub. L. No. 89-329, 79 Stat. 1219, codified as amended at 20 U.S.C. § 1090.

[104]26 U.S.C. § 6103(j).

[105]13 U.S.C. § 9(a)(3).

The Evidence-Based Policymaking Act of 2018, New Data Infrastructure, and Reforms to Existing Laws Could Improve Interagency Coordination

Forum participants discussed various opportunities to improve interagency collaboration (see table 15).

Table 15: Reported Opportunities to Improve Coordination Among Federal Statistical Agencies

Opportunity	Description
Build on the Evidence-Based Policymaking Act of 2018 (Evidence Act)	The Evidence Act lays the foundation for shared data services, and additional resources and regulations may strengthen the law's effect.
Leverage the leadership of Evidence Act officials	The Chief Data Officer and Statistical Official could use their authority to assist the coordination of data sharing agreements.
Move toward a shared data infrastructure	Developing secure platforms for interagency data sharing and analysis could support more effective coordination across agencies.
Reform existing laws	Modifying existing legislation could improve the statistical system's efficiency while protecting confidentiality.

Source: GAO analysis of the Federal Statistical System Forum. | GAO-25-107124

Build on the Evidence-Based Policymaking Act of 2018. Although passed in 2019, participants cited the Evidence-Based Policymaking Act of 2018 (Evidence Act) as a tool that could continue to help improve interagency collaboration. According to one participant, the Evidence Act established the foundation for shared frameworks, governance, and data services. Participants said this has potential to create greater consistency and collaboration—especially with respect to interagency data sharing—as the law is further implemented.[106] Participants said the law's effect may be contingent on additional resources, such as personnel, for implementation. Other participants cautioned against viewing the Evidence Act as a solution for eliminating all data sharing barriers, noting

[106]In 2023, we described how the Evidence Act promotes interagency coordination to facilitate evidence building. For example, the act establishes a Chief Data Officer Council to standardize data practices and encourage data sharing agreements among federal agencies. It also requires agencies to make their data available, upon request, to statistical agencies for the purpose of statistical production, with privacy safeguards. See GAO-23-105460.

that the Evidence Act promotes data sharing within the legal restrictions imposed by preexisting rules and laws.[107]

One Participant's View on Building on the Evidence Act

"The Evidence Act offers significant opportunities to advance modernization of the system. For example, two key areas within the Evidence Act—the regulatory framework and common processes/infrastructure—are game-changers for the [federal statistical system]."

Source: Participant in the Federal Statistical System Forum. | GAO-25-107124

Future regulations required by the Evidence Act may strengthen the law's impact. Participants noted that OMB's forthcoming regulations on the Presumption of Accessibility for Statistical Agencies and Units and Expanding Secure Access to Confidential Information Protection and Statistical Efficiency Act (CIPSEA) Data Assets will seek to clarify the Evidence Act's scope, enabling greater collaboration across the federal government.[108] The first regulation will enhance statistical agency access to data assets maintained by federal agencies for evidence building purposes by requiring a transparent process for statistical agencies to request such data. It will also establish clear and consistent standards, to the extent practicable, for complying with confidentiality and privacy laws. The second regulation will establish a process for Office of Management and Budget (OMB)-recognized statistical agencies and units, to the extent practicable, to expand sharing of CIPSEA data assets, while protecting the data from inappropriate access and use.

Leverage the leadership of Evidence Act officials. Participants discussed how leveraging the roles established by the Evidence Act—such as the Chief Data Officer and statistical official—could facilitate coordination among federal agencies to better meet evolving data

[107]For example, restrictions on sharing Internal Revenue Service data and Census data in 26 U.S.C. § 6103(j) and 13 U.S.C. § 9.

[108]This requirement is found in Title III of the Evidence Act, known as CIPSEA of 2018. See Pub. L. No. 115–435, 132 Stat. 5529, 5556 (2019), codified at 44 U.S.C. §§ 3581–3582.

needs.[109] According to one participant, the COVID-19 pandemic prompted the chief data officers of the Departments of Veterans Affairs and Health and Human Services to coordinate an agency-to-agency data sharing agreement, which the participant viewed to be successful. According to the participant, by standardizing and streamlining data sharing agreements, the Department of Veterans Affairs is obtaining Department of Health and Human Services data more efficiently than in the past, with one agreement covering multiple programmatic uses.

Move toward a shared data infrastructure. As previously discussed, while participants viewed shared services as a means of addressing the needs of diverse data user groups, they also said these services may enhance interagency collaboration by addressing critical barriers to data sharing and accessibility. For example, participants stated that the potential NSDS could facilitate secure data linkage among federal agencies—though another participant expressed skepticism that this would be a practical production tool for the U.S. Census Bureau, and potentially other statistical agencies, given the unique confidentiality, methodological, and legal requirements that govern statistical data. Participants also mentioned that expanded utilization of the Standard Application Process could generate further efficiencies, with one participant identifying it as a time-saving alternative to single use data sharing agreements.

In addition, one participant noted the potential for the Federal Statistical Research Data Centers to broaden the user base for statistical resources and to support critical decision-making by state and local governments. The participant highlighted the Centers' quick pivot to a virtual access model at the start of the COVID-19 pandemic, citing this as an instructive

[109]See Pub. L. No. 115-435, 132 Stat. 5529, 5531 & 5541 (2019); 5 U.S.C. § 314(a) and 44 U.S.C. § 3520(c). The Evidence Act established several new leadership positions for federal agencies, including the Chief Data Officer and the Statistical Official. The Chief Data Officer's responsibilities include managing the agency's data; ensuring the agency, to the extent practicable, maximizes its use of agency data to produce evidence; and serving as a member of the Chief Data Officer Council, which establishes government-wide best practices for the use, protection, dissemination, and generation of data, among other things. Statistical officials are designated officials with appropriate expertise, usually the head of a statistical agency or unit, who advise the parent agency on statistical policy, techniques, and procedures. For our recent reviews of the work of Chief Data Officers at various agencies, see GAO, *Chief Data Officer Council: Progress in Strengthening Federal Evidence-Based Policymaking*, GAO-23-105514 (Washington, D.C.: Dec. 15, 2022) and GAO, *Data Governance: Agencies Made Progress in Establishing Governance, but Need to Address Key Milestones*, GAO-21-152 (Washington, D.C.: Dec. 16, 2020).

example of how interagency coordination could expand tiered access to more and new categories of data users.

GAO Example of Challenges Agencies Face with Data Sharing

In 2022, we reported that the Internal Revenue Service does not collect certain demographic information about taxpayers—such as race, ethnicity, and sex—that other agencies, like the Census Bureau, do collect. However, legal restrictions under Titles 13 and 26 of the U.S. Code limit sharing of such data among agencies. We reported that this poses a barrier to analyzing the effects of tax policies in relation to taxpayers' demographics. To address this, we recommended Congress consider making legislative changes to facilitate interagency data sharing that would allow for more accurate, consistent, and systematic analyses of any effects of existing and proposed tax policies in relation to taxpayers' demographics in a secure manner that protects the confidentiality of those data.

Source: GAO, Tax Equity: Lack of Data Limits Ability to Analyze Effects of Tax Policy on Households by Demographic Characteristics, GAO-22-104553 (Washington, D.C.: May 2022). | GAO-25-107124

Reform existing laws. Participants stated that there are opportunities to modify existing legislation in ways that could improve the system's efficiency while protecting confidentiality. As noted previously, participants stated that competing laws and regulations (such as the tax code) may restrict sharing of administrative data among agencies due to agency-specific confidentiality requirements. These include current statutes that restrict the Internal Revenue Service and the Census Bureau from sharing key data with other agencies. Participants said that updating these laws and regulations, to allow agencies to share relevant data that statistical agencies do not directly collect, would help mitigate these challenges. For example, participants said that changes to the tax code, such as a change proposed by the President in 2021, would expand the Bureau of Economic Analysis' and the Bureau of Labor Statistics' access to certain, limited business data collected by the Internal Revenue Service, subject to restrictions for statistical purposes—thus yielding greater consistency between business registers and associated datasets.[110]

Participants offered mixed views on whether other laws and regulations should be revisited. To address concerns about confidentiality and data privacy, participants suggested refining existing legislation to clarify how statistical agencies access administrative data, including revisiting tax code and data sharing provisions that currently impose limitations. However, another participant flagged "very significant risks" to revisiting such legislation, noting that confidentiality protections were critical to having a high-quality census in 2020—given that it benefited from high response rates and leveraged sensitive administrative data from the Internal Revenue Service, among other agencies.

Concluding Summary

The discussion forum identified many strategies that participants said could improve the productivity, efficiency, and accountability of the federal statistical system. Alternative data sources may allow agencies to provide more granular, timely, and cost-effective products. Continued implementation of the Evidence Act, through the Trust Regulation and NSDS data sharing and analysis infrastructure, may allow agencies to share data more easily, widely, and securely for statistical purposes.

[110]U.S. Department of the Treasury, *General Expectations of the Administration's Fiscal Year 2022 Revenue Proposals* (Washington, D.C.: May 2021), https://home.treasury.gov/system/files/131/General-Explanations-FY2022.pdf.

Efforts to further understand user needs may allow agencies to provide a broader and more tailored range of data products, access, and services, and better target constrained budgets toward the programs that most effectively serve modern user needs. Although participants said these strategies have potential, many have not been attempted or fully implemented, and their effects remain unknown. Continued attention to these strategies will be needed in future years to fully understand their effects.

Agency Comments

We provided a draft of this report to OMB; the U.S. Departments of Agriculture, Commerce, Education, Health and Human Services, Homeland Security, Labor, and Veterans Affairs; and the U.S. National Science Foundation for review and comment. OMB, the U.S. Census Bureau within the U.S. Department of Commerce, and the U.S. Departments of Education and Labor provided us with technical comments, which we incorporated as appropriate. The Bureau of Economic Analysis within the U.S. Department of Commerce; the U.S. Departments of Agriculture, Commerce, Health and Human Services, Homeland Security, and Veterans Affairs; and the U.S. National Science Foundation did not have any comments on the report.

We are sending copies of this report to the appropriate congressional committees; the Director of OMB, Secretary of Agriculture, Secretary of Commerce, Secretary of Health and Human Services, Secretary of Labor, Secretary of Veterans Affairs, and the National Science Foundation Chief of Staff (performing the duties of the National Science Foundation Director); and other interested parties. In addition, the report is available at no charge on the GAO website at http://www.gao.gov.

If you or your staff have any questions about this report, please contact me at SmithJB@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in appendix IV.

//SIGNED//

Jared Smith
Director, Center for Statistics and Data Analysis, Applied Research and Methods

List of Committees

The Honorable Susan Collins
Chair
The Honorable Patty Murray
Vice Chair
Committee on Appropriations
United States Senate

The Honorable Lindsey Graham
Chairman
The Honorable Jeff Merkley
Ranking Member
Committee on the Budget
United States Senate

The Honorable Rand Paul, M.D.
Chairman
The Honorable Gary C. Peters
Ranking Member
Committee on Homeland Security and Governmental Affairs
United States Senate

The Honorable Tom Cole
Chairman
The Honorable Rosa DeLauro
Ranking Member
Committee on Appropriations
House of Representatives

The Honorable Jodey Arrington
Chairman
The Honorable Brendan Boyle
Ranking Member
Committee on the Budget
House of Representatives

The Honorable James Comer
Chairman
The Honorable Robert Garcia
Ranking Member
Committee on Oversight and Government Reform
House of Representatives

Appendix I: Forum Agenda

Wednesday, August 21, 2024

Agenda

10:30 am	Approach and Organization of the Event: Day 1
10:45 am	**Topic 1: Modernizing Federal Data Collection and Production**
12:15 pm	*Break for Lunch*
1:30 pm	**Topic 2: Public Use and Accessibility of Federal Statistical Products**
3:00 pm	*Break*
3:15 pm	**Topic 3: Resources, Productivity, Workforce, and Efficiency in a Modern Federal Statistical System**
4:45 pm	Closing of Day 1: Summary and Follow-Up Discussion
5:00 pm	*END OF DAY 1*

Thursday, August 22, 2024

10:00 am	Welcome and Organization of the Event: Day
10:15 am	**Topic 4: Innovation in Alternative Data Sources for Federal Statistics**
11:45 am	*Break for Lunch*
1:00 pm	**Topic 5: Enhancing Data-Sharing Across Federal Agencies**
2:30 pm	*Break*
2:45 pm	**Topic 6: Public Trust and Objectivity in Federal Statistics**
5:00 pm	*END OF DAY 2 AND ADJOURN*

Appendix II: List of Participants

	Affiliation at the Time of the Forum
Vipin Arora	Director, U.S. Bureau of Economic Analysis
Misha Belkindas	Managing Director, Open Data Watch
danah boyd	Partner Researcher, Microsoft Research
	Distinguished Visiting Professor, Georgetown University
Peggy G. Carr	Commissioner, National Center for Education Statistics, U.S. Department of Education
Matthew Crespi	Program Director, National League of Cities
John C. Haltiwanger	Distinguished University Professor, Department of Economics at the University of Maryland
Nicholas Hart	President, Data Foundation
Anna S. Hui	Director, Missouri Department of Labor and Industrial Relations
Ron S. Jarmin	Deputy Director and Chief Operating Officer, Census Bureau, U.S. Department of Commerce
Barry W. Johnson	Chief Data and Analytics Officer, Internal Revenue Service, U.S. Treasury Department
	Acting Director, Statistics of Income Division, Internal Revenue Service, U.S. Treasury Department
Sallie Ann Keller	Chief Scientist and Associate Director, Research and Methodology Directorate, Census Bureau, U.S. Department of Commerce
	Distinguished Professor in Biocomplexity, University of Virginia
Margaret C. Levenstein	Professor, School of Information, and Center Director, Inter-university Consortium for Political and Social Research, Institute for Social Research, University of Michigan
Felice J. Levine	Executive Director, American Educational Research Association
Jacob Malcom	Statistical Official and Director, Office of Policy Analysis, U.S. Department of the Interior
Bruce D. Meyer	McCormick Foundation Professor, Harris School of Public Policy, the University of Chicago
Mary Jo Mitchell	Director of Government Affairs, Population Association of America and Association of Population Centers

Brian C. Moyer	Director, National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services
Jeri Mulrow	Vice President and Director, Data Solutions Sector, Westat
Amy O'Hara	Research Professor, Massive Data Institute, University of Georgetown Director, Georgetown Federal Statistical Research Data Center
Karin Orvis	Chief Statistician of the United States, Office of Information and Regulatory Affairs, Office of Management and Budget
Steve Pierson	Director of Science Policy, American Statistical Association
Wendell Primus	Visiting Fellow, Center on Health Policy in the Economic Studies Program, Brookings Institution
Jerry Reiter	Professor of Statistical Science, Duke University
Emilda B. Rivers	Director, National Center for Science and Engineering Statistics, U.S. National Science Foundation
Spiro Stefanou	Administrator, Economic Research Service, U.S. Department of Agriculture
Markus Sovala	Director General, Statistics Finland
Kunal Talwar	Research Scientist, Apple
Eddie Thomas	Statistical Official and Director, National Center for Veterans Analysis, U.S. Department of Veterans Affairs
William (Bill) J. Wiatrowski	Deputy Commissioner, Bureau of Labor Statistics, U.S. Department of Labor

Appendix III: Objectives, Scope, and Methodology

This report examines forum participants' perspectives on factors that affect the statistical system's ability to (1) build and maintain public trust, (2) meet the needs of its users, (3) sustain and modernize its data collection, and (4) engage in effective interagency coordination.

This report is based primarily on a thematic analysis and summarization of the discussions that took place among expert and stakeholder participants during a forum convened on August 21-22, 2024, in Washington, D.C. The forum was planned and convened with the assistance of National Academies of Sciences, Engineering, and Medicine (National Academies) to better ensure that a breadth of expertise was brought to bear in its preparation. However, GAO was responsible for all final decisions regarding meeting substance and expert and stakeholder participation. Our methodology to convene the expert forum and produce this report included the following:

- developing an initial list of discussion topics through a systematic literature search for peer-reviewed articles, legislative materials, government reports and other materials;
- with assistance from the National Academies, developing a list of experts and stakeholders, and administering a questionnaire to inform the final selection of forum topics and participants;
- convening the forum with assistance from the National Academies and transcribing of the forum discussion;
- conducting a thematic analysis and summarization;
- conducting supplementary interviews to gain additional perspectives and expand on themes raised during the forum; and
- drafting the report based on consideration of the above and expert and stakeholder comments on the report draft.

Topic Development

To develop the forum's initial list of topics, we conducted a background (nonsystematic) literature search to develop a topical list of current and emerging issues in the federal statistical system. GAO librarians searched bibliographic databases and platforms including ProQuest and WorldCat for relevant peer reviewed materials, books, government and trade reports, and other relevant materials published within five years prior to and including September 2023. We identified 76 relevant articles.

For each identified article, a GAO statistician reviewed the abstract for relevance. Articles were selected for additional review if they directly pertained to the federal statistical system and statistical policy issues. We identified 51 articles as potentially relevant to our research questions. Our initial list identified challenges and opportunities facing the federal statistical system's (1) structure; (2) preparedness and robustness; and (3) data quality, relevance, and utility. We leveraged this list when selecting experts, stakeholders, and agenda topics for the forum.

Participant and Topic Selection

In coordination with the National Academies, we identified 29 experts and stakeholders to participate in the forum. For participant selection, we considered the experience and backgrounds of participants with the aim of convening a group of participants with diverse perspectives and relevant experience from (1) within the federal statistical system as an official within a Principal Statistical Agency, and (2) outside of the federal statistical system. We considered individuals to be within the federal statistical system if they were an agency official from the Office of Management and Budget's (OMB) Office of the Chief Statistician or from a Recognized Statistical Agency or Unit. We considered individuals outside of the federal statistical system if they were from (1) other federal agencies, (2) state and local governments, (3) research and academia, (4) the private sector, or (5) other national statistical offices and international organizations.

To identify potential stakeholders and experts to participate in the forum, the team developed a list of officials and representatives or organizations within the relevant groups listed above based on subject matter expertise, referrals from relevant experts, and experience, among other factors. Separately, the National Academies developed a list of potential participants based on a variety of criteria, including subject area expertise, professional affiliations and committee membership, publication history, and experience with prior consensus National Academies studies relevant to the federal statistical system. GAO and the National Academies worked together to combine the two lists to create a final list of 106 potential participants. The list of potential participants provided sufficient representation across subject areas for discussion and coverage across identified stakeholder groups, with expertise both within and outside of the federal statistical system.

To further identify individuals to participate in the forum and inform forum topic selection, we administered a web-based questionnaire to the 106

potential experts and stakeholders. We tested the survey with three statistical agency stakeholders from the Census Bureau, the Bureau of Labor Statistics, OMB, and one GAO stakeholder, and implemented changes to the survey based on their feedback. Potential participants were emailed the questionnaire by the National Academies in June and July 2024, and invited to complete the survey by July 15, 2024. Eighty-five of the 106 recipients completed the questionnaire by our required timeframe. If a respondent indicated interest in attending the forum but could not attend, we asked that they provide an alternate participant from their organization.

From questionnaire respondents, we created our invitee list based on a variety of factors, focusing on key areas of expertise and representation. We determined that the following experts and stakeholders must be represented:

- the Chief Statistician of the U.S., whose role is to coordinate the activities of the federal statistical system;
- participants from at least one principal statistical federal agency;
- participants from at least one federal agency that engages in statistical activities but is not a principal statistical agency; and
- participants from the following categories of expert and stakeholder groups: academia, federal, international, private, professional, and state or local government.

We used the following criteria to refine our invitee list:

- willingness to participate in the forum, based on response to our questionnaire;
- if willing to participate, availability to present on a topic as a roundtable focal point;
- balance and coverage across expert and stakeholder groups;
- balance and coverage among identified agenda topics for discussion;
- independence; and
- expert opinion from the National Academies.

Our selection methods ensured representation of varying perspectives across key groups, including 10 participants from the federal statistical system, two federal participants from outside the federal statistical

system, two from state or local government, one from a non-U.S. national statistical office, one from an international organization, five from research and academia, three from the private sector, and five from nonprofit professional organizations.

The questionnaire also prompted potential participants to select the top five issues they felt were the most important for the federal statistical system to address. In addition to GAO's review of the relevant literature, we used participant responses to this prompt to determine the forum topic areas.

Forum and Participant Follow-up

On August 21 – 22, 2024, we convened a 2-day, in-person forum comprising six sessions. Four participants were unable to attend in-person and accommodations were made for these participants to attend virtually. The forum topics were:

- modernizing federal data collection and production;
- public use and accessibility of federal statistical products;
- resources, productivity, workforce, and efficiency in a modern federal statistical system;
- innovation in alternative data sources for federal statistics;
- enhancing data sharing across federal agencies; and
- public trust and objectivity in federal statistics.

See appendix I for the forum agenda.

Each session was moderated by a GAO methodologist trained to guide discussions. The forum was structured as six guided roundtable discussions where four to six participants were invited to provide prepared opening remarks as a focal point and other participants were encouraged to openly comment on issues and respond to one another, although not all participants commented on all topics. Each roundtable session was 90 minutes long. A professional transcriptionist recorded audio and provided a verbatim transcript of the discussions to GAO.

We did not poll workshop participants or take votes on priority rankings of the various perspectives raised during the forum. Consequently, we do not provide counts or otherwise quantify the number of forum participants agreeing to a specific perspective. Further, since participants were

generating and discussing ideas as part of a free-flowing group discussion, the number of times a concept was (or was not) repeated does not necessarily indicate the level of consensus on that concept. Throughout the report, when summary statements are attributed to "participants" it indicates that more than one participant made a statement supporting a general point, and the views by others in support of the point were relatively uniform. In addition, written statements from expert panelists were received following the forum and were incorporated as supplementary content, as appropriate.

Thematic Analysis

To summarize the forum discussion, we reviewed the transcript and developed a process to thematically summarize the proceedings. Using an inductive approach, a team of four analysts iteratively developed a coding scheme to apply to the transcript by reviewing sections of the transcript and drafting a list of possible thematic codes. The codebook was refined through iterative rounds of code-application in which two analysts independently applied codes to the same selection of content, discussed differences in code application, and revised the codebook. The resulting codebook included three main thematic codes encompassing seven, six, and four-subcodes. After the codebook was finalized, an analyst was assigned to code each section of the transcript. For each participant statement that we used to support our summary of the forum, a second analyst reviewed the assigned codes and made their own determination. If code applications differed, the analysts discussed the differences and reached consensus on a final set of codes. We found that participants' statements coalesced into four key themes: building and maintaining public trust; meeting data user needs; engaging in effective interagency coordination; and sustainment and modernization of data collection.

Supplementary Interviews

After the forum, we conducted three interviews to gather more detail on non-federal and non-statistical agency use of statistical data, and on specific themes that emerged from thematic analysis of forum content. We conducted supplementary interviews with an agency official from the Department of the Interior who participated in the forum, an agency official from the Department of Homeland Security who did not participate in the forum, and a representative from state government who did not participate in the forum. We incorporated comments from the interviews,

as appropriate, and identified these sources separately from the forum discussions within our final report.

Forum Summary

We provided a draft summary of the report to participants and post-forum interviewees for their review and asked participants to assess and clarify quotes attributed to them in the forum transcript for accuracy. We incorporated their feedback as appropriate. To increase report clarity, there were instances where we made minor deletions or grammatical corrections to cited quotes, and asked participants to confirm clarified quotes. In addition, we used ellipses in cited quotes when we removed non-germane phrases to indicate alterations to reduce length or improve clarity.

While this report summarizes the key ideas that emerged during the forum, it is not intended to present an exhaustive catalogue of all ideas discussed by participants nor to represent all perspectives. In addition, the information presented in this summary does not necessarily represent the views of all participants, the organizations with which they are affiliated, or GAO.

We conducted our work from October 2023 to September 2025 in accordance with all sections of GAO's Quality Assurance Framework that are relevant to our objectives. The framework requires that we plan and perform the engagement to obtain sufficient and appropriate evidence to meet our stated objectives and to discuss any limitations in our work. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

Appendix IV: GAO Contact and Staff Acknowledgments

GAO Contact

Jared Smith, SmithJB@gao.gov

Staff Acknowledgments

In addition to the contact named above, Jeff Tessin (Assistant Director), Erin Villareal (Analyst-in-Charge), Leia Dickerson, Elizabeth Gooch, Scott Hiromoto, Amalia Konstas, Jason Rodriguez, Rebecca Sero, Norma-Jean Simon, Ashley Stewart, Frances Tirado, Walter Vance, Malika Williams, and Sirin Yaemsiri made key contributions to this report.