Data Silos, Dossiers, and Surveillance: The Unintended Risks of Federal Data Consolidation

April 28, 2025

Sharing data among federal databases — or consolidating those databases — may bring benefits but it also poses significant risks. On March 20, President Trump issued Executive Order 14243, "Stopping Waste, Fraud, and Abuse by Eliminating Information Silos," directing the consolidation of data across federal agencies. Although federal data sharing can bring benefits, the Executive Order also runs the risk of significant unintended consequences: surveillance, data misuse, and cybersecurity breaches; technical hurdles that will undermine government operations; and legal implications under several federal laws. Consolidated federal data would expand surveillance capabilities and exacerbate privacy risks under future Presidents, regardless of party. Instead, a better approach would be to build on pilot projects for purpose-driven data linkages established during the first Trump Administration.

DATA CONSOLIDATION: WHAT DOES THE EXECUTIVE ORDER DIRECT?

Expansion of data sharing and consolidation of federal databases is the Administration's official policy. President Trump's Executive Order directed agencies to share and "consolidate" federal records, nominally to combat fraud and waste. This goal, however, contradicts the law and Congress's intent as embodied in the Privacy Act of 1974 and other statutes; a president cannot use an executive order to undo a law enacted by Congress — a fact the Executive Order acknowledges in its final section. The role of the executive is to execute the laws that Congress passes. The Executive Order asserts three main goals:

- **Broad Sharing and "Consolidation" of Federal Records Across the Federal Government.** First, the Executive Order establishes sweeping access to our data held by federal agencies. It directs agencies to provide "Federal officials designated by the President or Agency Heads" with "full and prompt access to all unclassified agency records, data, software systems, and information technology systems." The Executive Order directs agencies to authorize and facilitate "both the intra- and inter-agency sharing and consolidation of unclassified agency records," including by taking steps to limit protections provided by the Privacy Act of 1974.

- **Targeting Unemployment Information.** Second, the Executive Order singles out the Department of Labor to "immediately" provide "designees" with "unfettered access to all unemployment data and related payment records." Although that provision seems targeted toward unemployment compensation claims, it could broadly sweep in data collected by the Bureau of Labor Statistics through surveys for the monthly "jobs report." Unemployment claims data contains incredibly sensitive information, including Social Security numbers, identification documents, wage records, employer history (including being fired or laid off), race, and disability.

- **Hoovering Up State Data.** Finally, the Executive Order directs agencies to "take all necessary steps" to ensure that the federal government has "unfettered access to comprehensive data from all State programs that receive Federal funding" — meaning data now held by state agencies for programs like the Supplemental Nutrition Assistance Program (SNAP), Medicaid, and K-12 public education will now be available to the federal government.

By Cody Venzke, Senior Policy Counsel, ACLU

DATA CONSOLIDATION CREATES RISKS OF SURVEILLANCE, DATA MISUSE, AND CYBERSECURITY BREACHES

A single, consolidated federal database would become a cudgel that, by design, breaks the protections meant to prevent governmental abuses of our data. Laws like the Privacy Act of 1974 were passed after President Nixon used tax data and personnel records to attack and embarrass political opponents. Those laws largely mandate that agencies share personal information only as necessary to carry out their work. Foundational privacy principles developed during the Nixon era similarly ensure that agencies are transparent in their use of personal information and that individuals consent to those uses. President Trump's Executive Order would undermine the very safeguards that Congress has established to prevent Watergate-era abuses from reoccurring. **Three threats are the most salient**:

First, data consolidation will enable **centralized dossiers** on everyone who interacts with federal or state governments — nearly everyone in the United States. Such centralized, consolidated databases would leap over the firewalls around agency data that prevent misuse and abuse. It would ultimately enable agencies across the federal government to search for personal information gathered into a single database. Federal officials would no longer be limited to the data collected for them to carry out their work but would instead have access to hundreds of data points on every individual.

For example, data consolidation would pave the way for biometrics collected as part of TSA PreCheck or Global Entry to be combined with or readily shared with the Federal Bureau of Investigation, the Internal Revenue Service, and others — permitting investigators to obtain facial scans for locating individuals, even for frivolous investigations. Likewise, firearm records held by federal firearms licensees, the FBI, the Bureau of Alcohol, Tobacco, Firearms and Explosives, or even state agencies could be accessed by other federal agencies, perhaps in determining suitability for federal benefits like Medicare or Social Security. Similarly, IRS record showing donations to civil rights organizations or Second Amendment groups could be used by political appointees in agencies to discredit perceived opponents. Centralized dossiers would enable social credit scores, akin to those used in authoritarian regimes, to vet "suitability and fitness" for federal employment, contracts, benefits, and more.

In many instances, this data sharing would run afoul of longstanding privacy protections, some of which are detailed below. Although the Executive Order directs agencies to act only "to the maximum extent consistent with law," that caveat may be too little protection, too late, as court decisions resolving legal questions may come only weeks or months after the data has been shared or consolidated. **Whatever data consolidation and surveillance are achieved now will be available for abuse by future Presidents of both parties.**

Second, consolidation of federal data opens the door to **misuses of data**. Misuse of data is not a new problem — just a few years ago, the Supreme Court heard a case in which a police officer agreed to search a state law enforcement database for a purported undercover agent in exchange for $5,000. In other instances, law enforcement and intelligence analysts have improperly accessed databases to snoop on former romantic partners or to share intimate photos discovered in the course of their duties.

Consolidation could exacerbate these abuses by expanding the data available to employees across the federal government. Firewalls around data exist largely to ensure that government employees have access only to the data they need to perform their duties. The Executive Order's directive that agencies provide "full and prompt access" to "all unclassified agency records" hops over those firewalls, and the Order goes even further in directing agencies to "rescind or modify all agency guidance that serves as a barrier to the inter- or intra-agency sharing." The Order makes little room for safeguards to prevent misuse of data within the federal government.

And third, consolidating federal data also increases **the risks of cybersecurity incidents**. A single, consolidated database would be attractive to foreign adversaries, who have previously targeted data held by the Office of Personnel Management, large school districts, communications networks, and communications providers, affecting tens of thousands of businesses and millions of people. Likewise, rapid consolidation may make it difficult to establish robust cybersecurity protections. Federal law mandates that information systems meet minimum security controls, but agencies have long faced obstacles in meeting those requirements. The Government Accountability Office recently stated that agencies have "varied in their efforts to implement key security practices for cloud services, which provide on-demand access to shared resources such as networks, servers, and data storage." Cloud services would be essential for consolidation of federal records, but the necessary cybersecurity measures are lagging behind. Rapid consolidation could lead to cybersecurity missteps, with potentially disastrous results.

DATA CONSOLIDATION FACES SIGNIFICANT TECHNICAL OBSTACLES THAT MAY LEAD TO SERIOUS HARMS

In addition to posing risks of surveillance and data misuse, data consolidation faces significant technical obstacles that will inevitably result in inaccuracies that delay or deny benefits for constituents, wrongly flag Americans for investigation, or waste taxpayer resources. As agencies carry out President Trump's Executive Order to authorize "sharing and consolidation" of federal data, they will need to ensure that the data they are consolidating belongs together — that the John Henry Doe in *this* IRS record is the same as John H. Doe in *this* SSA record, and that both are the same as "John Doe" in *this* state Medicaid record. Consolidating those records is much harder and messier than one might expect, as databases may be in different formats, have been created using different coding languages, have incorrect, incomplete, or outdated information, or simply be too vague to be definitively connected.

Matching records is so vexing that it has given rise to an entire branch of data science known as "entity resolution," and it's no surprise that previous efforts to consolidate or match data have suffered from inaccuracies. For example:

- **Student Aid and IRS Data Matching:** In 2024, the Department of Education (ED) encountered numerous issues that hampered its efforts to match federal student aid data with data from the IRS. For approximately 15 percent of FAFSA applications submitted that year, the IRS's Direct Data Exchange tool reported incorrect income information, affecting the amount of financial support awarded to students. Moreover, for a set of students whose families had updated their tax returns after filing, the IRS retrieval tools inconsistently reported updated information in some instances and the original information in others.

- **Matching Across State SNAP Databases:** The Department of Health and Human Services (HHS) facilitates a data-matching program across state SNAP programs to help identify individuals claiming benefits in multiple states. The program, however, has been plagued by operational challenges. In 2017, the Government Accountability Office (GAO) surveyed the directors of all 51 state SNAP programs. The directors reported that data often was not sufficiently detailed or recent enough to permit effective matching or was collected for purposes not suited to SNAP's eligibility requirements. A new matching system is currently being launched but is not expected to be implemented until the fall of 2027.

- **Errors in State-Level Anti-Fraud Efforts:** States have long engaged in other data matching and consolidation efforts to address fraud and waste. Those efforts have likewise been hampered by errors. For example, California's unemployment insurance system froze or delayed support for millions of individuals, often due to faulty data matching: people were wrongly disenrolled due to typos in databases, inconsistent hyphenation, nicknames, the use of initials, incorrect dates, and even names that were too long to fit into the databases. Similarly, Tennessee wrongly terminated Medicaid support for thousands of families, often because conversion to a new eligibility system erroneously placed recipients in an administrative category that required further review and delayed benefits.

The experience with existing efforts to match and consolidate databases urges caution for the more extensive effort envisioned by President Trump's "information silos" Executive Order. Recent efforts by the Administration underscore the potential risks; for example, after accessing data at SSA, the Department of Governmental Efficiency purportedly discovered 120-year-olds collecting benefits — a conclusion that arose from a lack of familiarity with the older codebase used for SSA's databases and how it calculates dates for fields with missing data. Efforts are now underway to quickly re-write the entirety of SSA's code base, and efforts to conduct identity proofs have already caused two system outages.

Despite the risks of unintended consequences, federal agencies are already consolidating federal data, including by relying on AI — systems that are often untested and, in early stage of AI development, may lead to inaccurate, unsafe, or unfair results. For example:

- An official with the General Services Administration (GSA) recently stated that the federal government needs a centralized data repository to achieve its "AI-first strategy," and that privacy laws are merely a "roadblock" to accomplishing that goal.

- Meanwhile, other agencies are "working to authorize the use of AI tools," and AI has been deployed at ED, with access to grant and financial information, resulting in "a massive firehose of data being sent to [an] AI company's servers."

- Discussions are similarly underway at the IRS to tap data analytics and AI company Palantir to build a portal to make highly protected IRS data available across the federal government.

- AI has likewise been deployed to analyze the "five bullet point" emails that were mandated from federal employees to summarize their weekly activity, to assess federal employees for termination, and to advise federal employees at GSA.

The fast deployment of AI throughout critical governmental functions raises serious questions about whether its use complies with the Administration's recent guidance ensuring that AI is vetted and trustworthy *before* it is used by federal agencies. If not, the rapid deployment of AI could result in the cutting off funding for critical local programs, improper termination of federal employees, or AI that dispenses incorrect advice to Americans and federal workers.

DATA CONSOLIDATION WITHOUT SAFEGUARDS RAISES SIGNIFICANT LEGAL QUESTIONS

The Executive Order raises significant concerns about its compliance with legal restrictions on federal and state data. For example:

- **The Privacy Act of 1974**, 5 U.S.C. § 552a, which has figured prominently in recent litigation, prohibits disclosure of records from any federal agency's "system of records," including to other agencies. The law includes a variety of exceptions, such as disclosures to agency employees for "performance of their duties" and for "routine uses" that are compatible with the original purpose of collection and published in the Federal Register. In ongoing litigation, courts have enjoined data access at a variety of agencies. For example, one court rejected the argument that federal employees need unfettered data access to combat fraud, stating:

  > "[The government] never identified or articulated even a single reason for which [it] needs unlimited access to SSA's entire record systems, thereby exposing personal, confidential, sensitive, and private information that millions of Americans entrusted to their government. Indeed, the government has not even attempted to explain why a more tailored, measured, titrated approach is not suitable to the task. Instead, the

> government simply repeats its incantation of a need to modernize the system and uncover fraud. Its method of doing so is tantamount to hitting a fly with a sledgehammer."

The Privacy Act also permits individuals to sue and recover monetary damages.

- **Internal Revenue Code,** 26 U.S.C. § 6103, states that no federal or state employee "shall disclose any [tax] return or return information." It includes several narrowly drawn exceptions such as disclosure to chairs of the House Ways & Mean and Senate Finance Committees, the President, or any agency head upon request of the President. Violations may be subject to both criminal and civil penalties.

- The **E-Government Act of 2002**, 44 U.S.C. § 3501 note, requires Privacy Impact Assessments (PIA) prior to developing or deploying information technology that collects, maintains, or distributes information in an "identifiable form" or initiating an information collection from 10 or more members of the public. PIAs must describe the information collected, its purpose and uses, and how it will be shared and secured.

- The **Confidential Information Protection and Statistical Efficiency Act (CIPSEA)**, 44 U.S.C. § 3572, requires that personal information collected for statistical purposes by the 16 Recognized Statistical Agencies and Units such as the Census Bureau and the Bureau of Labor Statistics be used only for statistical purposes and not be disclosed without consent.

- Likewise, dozens of federal laws protect individual information in specific sectors. For example, the **Family Educational Rights and Privacy Act (FERPA)**, 20 U.S.C. § 1232g and 34 C.F.R. § 99.31, restricts public schools and school districts from sharing student's education records, including with law enforcement or the federal government. Similarly, the **Health Insurance Portability and Accountability Act (HIPAA)** 45 C.F.R. §§ 164.500–.526, restricts certain healthcare providers and other entities, including Medicare and Medicaid, from disclosing protected health information, including to law enforcement and the federal government.

- Finally, several federal laws specifically prohibit the federal government from building national databases on certain kinds of individuals. For example, the **Elementary and Secondary Education Act**, 20 U.S.C. § 7911, and the **Higher Education Act**, 20 U.S.C. § 1015c, both disclaim providing *any* authority to the federal government to develop a nationwide database of personally identifiable information on students, with very limited exceptions.

POLICY APPROACHES FOR EFFECTIVE, SAFE DATA LINKAGES

To avoid these legal ramifications, the Administration should work to create safe, effective data linkages that preserve privacy protections while enabling data use — a long-standing goal of policies passed by Congress and enacted by the first Trump Administration. Those policies have eschewed a single federal database in favor of temporary, purpose-drive linkages of data.

One example is the bipartisan "National Secure Data Service" (NSDS), developed during the first Trump Administration. A commission under President Trump examined possibilities for establishing secure, private, and effective data sharing; it rejected the idea of a massive, unified "clearinghouse" of federal data.

Instead, the commission concluded:

> “[Federal data sharing] should be designed to link data on an individual project basis only. In contrast to a clearinghouse, it should not lead to the establishment of a store of data that grows with every research project conducted. The data linked for a project through the NSDS should be kept structurally separate from other data linked through the NSDS for other projects. By strictly enforcing this design, the NSDS will further the goal of increased access to and use of data for specific research and evaluation efforts, without unduly increasing the potential for privacy harm.”

Those structural protections would be accompanied by other privacy protections like deleting data and closing linkages once they are no longer needed, removing identifiers, establishing a public steering committee, and creating a public inventory of approved linkages. **The “information silos” Executive Order lacks those protections and could lead to serious unintended harm.**

A pilot program to further develop data linkages is already underway at the National Science Foundation and would be a better route for the Trump Administration and Congress than widespread consolidation.

CONCLUSION

Combining data across federal agencies carries significant risks — it might be used to create centralized dossiers on Americans, empower federal surveillance, curtail state autonomy, and increase the risk of data misuse and cybersecurity breaches. Although data sharing can make programs more efficient and more effective, broad unfettered sharing of federal data carries risks that outweigh those benefits. Instead, the Administration should continue to develop safe, secure data linkages to ensure that federal data sharing is privacy protective and purposeful.