FAS FEDERATION OF AMERICAN SCIENTISTS

Richard L. Revesz
Administrator, Office of Information and Regulatory Affairs
Office of Management and Budget
725 17th St NW, Washington, DC 20503

December 16, 2024

RE: Request for Information on the Executive Branch Agency Handling of Commercially Available Information Containing Personally Identifiable Information, 2024-23773 (89 FR 83517)

To Whom it May Concern,

The Federation of American Scientists (FAS) is a non-partisan, nonprofit organization committed to using science and technology to benefit humanity by delivering on the promise of equitable and impactful policy. FAS believes that society benefits from a federal government that harnesses science, technology, and innovation to meet ambitious policy goals and deliver impactful results to the public.

We are writing in response to your Request for Information on the Executive Branch Agency Handling of Commercially Available Information (CAI) Containing Personally Identifiable Information (PII). Specifically, we will be answering questions 2 and 5 in your request for information:

> *2. What frameworks, models, or best practices should [the White House Office of Management and Budget] consider as it evaluates agency standards and procedures associated with the handling of CAI containing PII and considers potential guidance to agencies on ways to mitigate privacy risks from agencies' handling of CAI containing PII?*
>
> *5. Agencies provide transparency into the handling of PII through various means (e.g., policies and directives, Privacy Act statements and other privacy notices at the point of collection, Privacy Act system of records notices, and privacy impact assessments). What, if any, improvements would enhance the public's understanding of how agencies handle CAI containing PII?*

Background

In the digital landscape, commercially available information (CAI) represents a vast ecosystem of personal data that can be easily obtained, sold, or licensed to various entities. The Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (EO 14110) defines CAI comprehensively as information about individuals or groups that is publicly accessible, encompassing details like device information and location data.

A 2017 report by the Georgetown Law Review found that 63% of Americans can be uniquely identified using just three basic attributes—gender, birth date, and ZIP code—with an astonishing 99.98% of individuals potentially re-identifiable from a dataset containing only 15 fundamental characteristics. This vulnerability underscores the critical challenges of data privacy in an increasingly interconnected world.

CAI takes on heightened significance in the context of artificial intelligence (AI) deployment, as these systems enable both data collection and the use of advanced inference models to analyze datasets and produce predictions, insights, and assumptions that reveal patterns or relationships not directly evident in the data. Some AI systems can allow the intentional or unintentional reidentification of supposedly anonymized private data. These capabilities raise questions about privacy, consent, and the potential for unprecedented levels of personal information aggregation and analysis, challenging existing data protection frameworks and individual rights.

The United States federal government is one of the largest customers of commercial data brokers. Government entities increasingly use CAI to empower public programs, enabling federal agencies to augment decision-making, policy development, and resource allocation and enrich research and innovation goals with large yet granular datasets. For example, the National Institutes of Health have discussed within their data strategies how to incorporate commercially available data into research projects. The use of commercially available electronic health records is essential for understanding social inequalities within the healthcare system but includes sensitive personal data that must be protected.

However, government agencies face significant public scrutiny over their use of CAI in areas including law enforcement, homeland security, immigration, and tax administration. This scrutiny stems from concerns about privacy violations, algorithmic bias, and the risks of invasive surveillance, profiling, and discriminatory enforcement practices that could disproportionately harm vulnerable populations. For example, federal agencies like Immigration and Customs Enforcement (ICE) and Customs and Border Protection (CBP) have used broker-purchased location data to track individuals without warrants, raising constitutional concerns.

In 2020, the American Civil Liberties Union filed a Freedom of Information Act lawsuit against several Department of Homeland Security (DHS) agencies, arguing that the DHS's use of cellphone data and data from smartphone apps constitutes unreasonable searches without a warrant and violates the Fourth Amendment. A report by the Electronic Frontier Foundation found that CAI was used for mass surveillance practices, including geofence warrants that query all phones in specific locations, further challenging constitutional protections.

While the Privacy Act of 1974 covers the use of federally collected personal information by agencies, there is no explicit guidance governing federal use of third-party data. The bipartisan Fourth Amendment is Not for Sale Act (H.R.4639) would bar certain technology providers—such as remote computing service and electronic communication service providers—from sharing the contents of stored electronic communications with anyone (including government actors) and from sharing customer records with government agencies. The bill has passed the House of Representatives in the 118th Congress but has yet to pass the Senate as of December 2024. Without protections in statute, it is imperative that the federal government crafts clear guidance on the use of CAI containing PII in AI systems. In this response to the Office of Management and Budget's (OMB) request for information, FAS will outline three policy ideas that can improve how federal agencies navigate the use of CAI containing PII, including in AI use.

Summary of Recommendations

The federal government is responsible for ensuring the safety and privacy of the processing of personally identifiable information within commercially available information used for the development and deployment of artificial intelligence systems. For this RFI, FAS brings three proposals to increase government capacity in ensuring transparency and risk mitigation in how CAI containing PII is used, including in agency use of AI:

1. **Enable FedRAMP to Create an Authorization System for Third-Party Data Sources:** An authorization framework for CAI containing PII would ensure a standardized approach for data collection, management, and contracting, mitigating risks, and ensuring ethical data use.
2. **Expand Existing Privacy Impact Assessments (PIA) to Incorporate Additional Requirements and Periodic Evaluations:** Regular public reports on CAI sources and usage will enable stakeholders to monitor federal data practices effectively.
3. **Build Government Capacity for the Use of Privacy Enhancing Technologies to Bolster Anonymization Techniques** by harnessing existing resources such as the United States Digital Service (USDS).

Recommendation 1: Enable FedRAMP to Create an Authorization System for Third-Party Data Sources

Government agencies utilizing CAI should implement a pre-evaluation process before acquiring large datasets to ensure privacy and security. OMB, along with other agencies that are a part of the governing board of the Federal Risk and Authorization Management Program (FedRAMP), should direct FedRAMP to create an authorization framework for third-party data sources that contract with government agencies, especially data brokers that provide CAI with PII, to ensure that these vendors comply with privacy and security requirements. FedRAMP is uniquely positioned for this task because of its previous mandate to ensure the safety of cloud service providers used by the federal government and its recent expansion of this mandate to standardize AI technologies. The program could additionally harmonize its new CAI requirements with its forthcoming AI authorization framework.

When designing the content of the CAI authorization, a useful benchmark in terms of evaluation criteria is the Ag Data Transparent (ADT) certification process. Companies applying for this certification must submit contracts and respond to 11 data collection, usage, and sharing questions. Like the FedRAMP authorization process, a third-party administrator reviews these materials for consistency, granting the ADT seal only if the company's practices align with its contracts. Any discrepancies must be corrected, promoting transparency and protecting farmers' data rights. The ADT is a voluntary certification, and therefore does not provide a good model for enforcement. However, it does provide a framework for the kind of documentation that should be required. The CAI authorization should thus include the following information required by the ADT certification process:

- **Data source:** The origin or provider of the data, such as a specific individual, organization, database, device, or system, that supplies information for analysis or processing, as well as the technologies, platforms, or applications used to collect data. For example, the authorization framework should identify if an **AI system collected, compiled, or aggregated a CAI dataset.**
- **Data categories:** The classification of data based on its format or nature, such as structured (e.g., spreadsheets), unstructured (e.g., text or images), personal (e.g., names, Social Security numbers), or non-personal (e.g., aggregated statistics).
- **Data ownership:** A description of any agreements in place that define which individual or organization owns the data and what happens when that ownership is transferred.
- **Third-party data collection contractors:** An explanation of whether or not partners or contractors associated with the vendor have to follow the company's data governance standards.
- **Consent and authorization to sell to third-party contractors:** A description of whether or not there is an explicit agreement between data subjects (e.g., an individual using an application) that their data can be collected and sold to the government or another entity for different purposes, such as use to train or deploy an AI system. In addition, a description of the consent that has been obtained for that use.
- **Opt out and deletion:** Whether or not the data can be deleted at the request of a data subject, or if the data subject opt out of certain data use. A description of the existing mechanisms where individuals can decline or withdraw consent for their data to be

collected, processed, or used, ensuring they retain control over their personal information.

- **Security safeguards and breach notifications:** The measures and protocols implemented to protect data from unauthorized access, breaches, and misuse. These include encryption, access controls, secure storage, vulnerability testing, and compliance with industry security standards.

Unlike the ADT, a FedRAMP authorization process can be strictly enforced. FedRAMP is mandatory for all cloud service providers working with the executive branch and follows a detailed authorization process with evaluations and third-party auditors. It would be valuable to bring that assessment rigor to federal agency use of CAI, and would help provide clarity to commercial vendors.

The authorization framework should also document the following specific protocols for the use of CAI within AI systems:

- Provide a detailed explanation of which datasets were aggregated and the efforts to minimize data. According to a report by the Information Systems Audit and Control Association (ISACA), singular data points, when combined, can compromise anonymity, especially when placed through an AI system with inference capabilities.
- Type of de-identification or anonymization technique used. Providing this information helps agencies assess whether additional measures are necessary, particularly when using AI systems capable of recognizing patterns that could re-identify individuals.

By setting these standards, this authorization could help agencies understand privacy risks and ensure the reliability of CAI data vendors before deploying purchased datasets within AI systems or other information systems, therefore setting them up to create appropriate mitigation strategies.

By encouraging data brokers to follow best practices, this recommendation would allow agencies to focus on authorized datasets that meet privacy and security standards. Public availability of this information could drive market-wide improvements in data governance and elevate trust in responsible data usage. This approach would support ethical data governance in AI projects and create a more transparent, publicly accountable framework for CAI use in government.

Recommendation 2: Expand Privacy Impact Assessments (PIA) to Incorporate Additional Requirements and Periodic Evaluations

Public transparency regarding the origins and details of government-acquired CAI containing PII is critical, especially given the largely unregulated nature of the data broker industry at the federal level. Privacy Impact Assessments (PIAs) are mandated under Section 208 of the 2002

E-Government Act and OMB Memo M-03-22, and can serve as a vital policy tool for ensuring such transparency. Agencies must complete PIAs at the outset of any new electronic information collection process that includes “information in identifiable form for ten or more persons.” Under direction from Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, OMB issued a request for information in April 2024 to explore updating PIA guidance for AI-era privacy concerns, although new guidance has not yet been issued.

To ensure that PIAs can effectively provide transparency into government practices on CAI that contains PII, **we recommend that OMB provide updated guidance requiring agencies to regularly review and update their PIAs at least every three years, and also require agencies to report more comprehensive information in PIAs.** We provide more details on these recommendations below.

First, OMB should guide agencies to periodically update their PIAs to ensure evolutions in agency data practices are publicly captured, which is increasingly important as data-driven AI systems are adopted by government actors and create novel privacy concerns. Under OMB Memo M-03-22, agencies must initiate or update PIAs when new privacy risks or factors emerge that affect the collection and handling of PII, including when agencies incorporate PII obtained from commercial or public sources into existing information systems. However, a public comment submitted by the Electronic Privacy Information Center (EPIC) pointed out that many agencies fail to publish and update required PIAs in a timely manner, indicating that a stricter schedule is needed to maintain accountability for PIA reporting requirements. As data privacy risks evolve through the advancement of AI systems, increased cybersecurity risks, and new legislation, it is essential that a minimum standard schedule for updating PIAs is created to ensure agencies provide the public with an up-to-date understanding of the potential risks resulting from using CAI that includes PII. For example, the European Union’s General Data Protection Regulation (Art. 35) requires PIAs to be reconducted every three years.

Second, agency PIAs should report more detailed information on the CAI’s source, vendor information, contract agreements, and licensing arrangements. A frequent critique of existing PIAs is that they contain too little information to inform the public of relevant privacy harms. Such a lack of transparency risks damaging public trust in government. One model for expanded reporting frameworks for CAI containing PII is the May 2024 Policy Framework for CAI, established for the Intelligence Community (IC) by the Office of the Director of National Intelligence (ODNI). This framework requires the IC to document and report “the source of the Sensitive CAI and from whom the Sensitive CAI was accessed or collected” and “any licensing agreements and/or contract restrictions applicable to the Sensitive CAI”. OMB should incorporate these reporting practices into agency PIA requirements and explicitly require agencies to identify the CAI data vendor in order to provide insight into the source and quality of purchased data.

Many of these elements are also present in Recommendation 1, for a new FedRAMP authorization framework. However, that recommendation does not include existing agency projects using CAI or agencies that could contract CAI datasets outside of the FedRAMP authorization. Including this information within the PIA framework also allows for an iterative understanding of privacy risks throughout the lifecycle of a project using CAI.

By obligating agencies to provide more frequent PIA updates and include additional details on the source, vendor, contract and licensing arrangements for CAI containing PII, the public gains valuable insight into how government agencies acquire, use, and manage sensitive data. These updates to PIAs would allow civil society groups, journalists, and other external stakeholders to track government data management practices over time during this critical juncture where federal uptake of AI systems is rapidly increasing.

Recommendation 3: Build Government Capacity for the Use of Privacy Enhancing Technologies to Bolster Anonymization Techniques

Privacy Enhancing Technologies (PETs) are a diverse set of tools that can be used throughout the data lifecycle to ensure privacy by design. They can also be powerful tools in ensuring that PII within CAI) is adequately anonymized and secure. OMB should collect information on current agency PET usage, gather best practices, and identify deployment gaps. To address these gaps, OMB should collaborate with agencies like the USDS to establish capacity-building programs, leveraging initiatives like the proposed "Responsible Data Sharing Core" to provide expert consultations and enhance responsible data-sharing practices.

Meta’s Open Loop project identified eight types of PETs that are ripe to be deployed in AI systems, categorizing them into maturity levels, context of deployment, and limitations. One type of PET is differential privacy, a mathematical framework designed to protect individuals' privacy in datasets by introducing controlled noise to the data. This ensures that the output of data analysis or AI models does not reveal whether a specific individual’s information is included in the dataset. The noise is calibrated to balance privacy with data utility, allowing meaningful insights to be derived without compromising personal information. Differential privacy is particularly useful in AI models that rely on large-scale data for training, as it prevents the inadvertent exposure of PII during the learning process. Within the federal government, the U.S. Census Bureau is using differential privacy to anonymize data while preserving its aggregate utility, ensuring compliance with privacy regulations and reducing re-identification within datasets.

Scaling the use of PETs in other agencies has been referenced in several U.S. government strategy documents, such as the National Strategy to Advance Privacy-Preserving Data Sharing and Analytics, which encourages federal agencies to adopt and invest in the development of PETs, and the Executive Order (EO) on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, which calls for federal agencies to identify where they could use PETs. As a continuation of this EO, the National Science Foundation and the Department of Energy established a Research Coordination Network on PETs that will "address the barriers to widespread adoption of PETs, including regulatory considerations."

Although the ongoing research and development of PETS is vital to this growing field, there is an increasing need to ensure these technologies are implemented across the federal government. To kick this off, OMB should collect detailed information on how agencies currently use PETs, especially in projects that use CAI containing PII. This effort should include gathering best practices from agencies with successful PET implementations, such as the previous U.S. Census Bureau's use of differential privacy. Additionally, OMB should identify gaps in PET deployment, assessing barriers such as technical capacity, funding, and awareness of relevant PETs. To address these gaps, OMB should collaborate with other federal agencies to design and implement capacity-building programs, equipping personnel with the knowledge and tools needed to integrate PETs effectively. For example, a forthcoming FAS' Day One Project publication, "Increasing Responsible Data Sharing Capacity throughout Government," seeks to harness existing government capabilities to build government capacity in deploying PETs. This proposal aims to enhance responsible data sharing in government by creating a capacity-building initiative called the "Responsible Data Sharing Core" (RDSC).[1] Managed by the USDS, the RDSC would deploy fellows and industry experts to agencies to consult on data use and sharing decisions and offer consultations on which PETs are appropriate for different contexts.

Conclusion

The federal government's increasing reliance on CAI containing PII presents significant privacy challenges. The current landscape of data procurement and AI deployment by agencies like ICE, CBP, and others raises critical concerns about potential Fourth Amendment violations, discriminatory profiling, and lack of transparency.

The ideas proposed in this memo—implementing FedRAMPamp authorization for data brokers, expanding privacy impact assessment requirements, and developing capacity-building programs for privacy-enhancing technologies—represent crucial first steps in addressing these systemic risks. As AI systems become increasingly integrated into government processes, maintaining a delicate balance between technological advancement and fundamental

[1] Forthcoming. Increasing Responsible Data Sharing Capacity throughout Government by Rachel Cummings, Shlomi Hod, Palak Jain, Gabriel Kaptchuk, Tamalika Mukherjee, Priyanka Nanayakkara, Jayshree Sarathy, and Jeremy Seeman

constitutional protections will be paramount to preserving individual privacy, promoting responsible adoption, and maintaining public trust.

We appreciate the opportunity to contribute to this Request for Information on Executive Branch Agency Handling of Commercially Available Information Containing Personally Identifiable Information. Please contact clangevin@fas.org if you have any questions or need additional information.

Sincerely,

Clara Langevin
AI Policy Specialist
Federation of American Scientists

Karinna Gerhardt
Manager, Emerging Technologies and Competitiveness
Federation of American Scientists