CENTER FOR DEMOCRACY & TECHNOLOGY

December 16, 2024

To: Kevin Herms
Office of Management and Budget
MBX.OMB.CAI_RFI_FY24@omb.eop.gov

Submitted via regulations.gov

Re: Request for Information: Executive Branch Agency Handling of Commercially Available Information Containing Personally Identifiable Information, OMB-2024-0007-0001

The Center for Democracy & Technology (CDT) respectfully submits these comments in response to the Office of Management and Budget (OMB) Request for Information on executive branch agency handling of commercially available information (CAI) containing personally identifiable information (PII). CDT is a nonprofit 501(c)(3) organization that works to advance civil rights and civil liberties in the digital age. CDT's work includes advocating for responsible government procurement and use of technology, including AI, to ensure that government services and benefits are delivered efficiently while protecting people's privacy and civil rights. We also advocate for reasonable checks and balances on the ability of the government — and in particular the government's law enforcement and intelligence components — to access, collect, and store individuals' data to ensure all people can seek information and express themselves freely.

The limited use of CAI by public agencies can potentially play a legitimate role in government operations, but it comes with significant privacy, financial, and ethical risks and drawbacks. If public agencies utilize CAI, two overarching principles should be carried front-of-mind by government officials:

1. ***Public agencies should avoid paying for CAI containing PII obtained from a data broker***

Many data broker practices have proven harmful to individuals by undermining their privacy, safety, rights, and wellbeing. The proliferation of this business model has eroded people's ability to control their own information, to decide for themselves when and how their information is

used, and in many cases to ensure that information purported to describe them is correct.[1] In fact, information attained by data brokers like someone's full name, phone number, close relatives, and/or home address can be sold to individuals with nefarious intentions to enable stalking, abuse, and harassment.[2] Government agencies should avoid supporting the existence of this data ecosystem by minimizing their purchases of PII, even if this information could play a constructive role in improving public service delivery.

2. *Law enforcement and intelligence agencies should not be allowed to purchase CAI that would otherwise necessitate a warrant, court order, or other legal process to obtain*

Warrant requirements and independent court approval for surveillance are a pillar of our democracy, and law enforcement and intelligence agencies' purchase of CAI undermines this foundational protection. Strict rules should limit commercial acquisition of Americans' data by law enforcement and intelligence agencies.[3] Such purchases circumvent basic statutory and Fourth Amendment rights and safeguards against abuse. When law enforcement and intelligence agencies purchase Americans' private data, they can bypass legal requirements that act as a critical safeguard. Permitting these purchases also risks bulk collection absent any suspicion of wrongdoing, a form of dragnet surveillance that has always been anathema to American values and was firmly rebuked by Congress within the past decade.[4]

Purchasing data such as cell phone location information, web browsing activities, and communications metadata can give police an intimate look into individuals' lives; it may reveal their political and religious beliefs, their medical conditions and treatments; or their personal and romantic associations.[5] Such information could be weaponized for selective prosecutions,

---

[1] Ridhi Shetty, *CDT Comments to CFPB Lay Out Data Broker Harms That Should Be Held Accountable*, Center for Democracy & Technology (Jul 26, 2023) https://cdt.org/insights/cdt-comments-to-cfpb-lay-out-data-broker-harms-that-should-be-held-accountable/ [https://perma.cc/P2JH-EGEC].

[2] Joseph Cox, *T-Mobile 'Put My Life in Danger' Says Woman Stalked with Black Market Location Data*, Vice (Aug 21, 2019) https://www.vice.com/en/article/t-mobile-put-my-life-in-danger-says-victim-of-black-market-location-data/ [https://perma.cc/RAV6-9H9T].

[3] Sharon Bradford Franklin, Greg Nojeim, Dhanaraj Thakur, *Report – Legal Loopholes and Data for Dollars: How Law Enforcement and Intelligence Agencies Are Buying Your Data from Brokers*, Center for Democracy & Technology (Dec 9, 2021) https://cdt.org/insights/report-legal-loopholes-and-data-for-dollars-how-law-enforcement-and-intelligence-agencies-are-buying-your-data-from-brokers/ [https://perma.cc/NR7D-QNVJ].

[4] Center for Democracy & Technology, *Victory: Passage of USA FREEDOM Act Reins in NSA Surveillance* (Jun 2, 2015) https://cdt.org/press/victory-passage-of-usa-freedom-act-reins-in-nsa-surveillance/ [https://perma.cc/D2X5-F4L6].

[5] See e.g.: Al Jazeera, *US military buys location data of popular Muslim apps: Report* (Nov 17, 2020) https://www.aljazeera.com/news/2020/11/17/report-us-military-buying-location-data-on-popular-muslim-apps [https://perma.cc/UR8Y-6NCT]; Alexandra Reeve Givens & Jocelyn Frye, *Protect Reproductive Privacy: Close the Health Data Loophole*, Tech Policy Press (Sept 9, 2024) https://www.techpolicy.press/protect-reproductive-privacy-close-the-health-data-loophole/ [https://perma.cc/K26E-WBGL].

released to attack dissidents, or stockpiled to chill free expression, risks that grow larger with the scale of collection.

Court oversight and approval rules act as a critical check against the dangers of unfettered surveillance, and should bind law enforcement and intelligence agency acquisition, even when data is commercially available to private parties. Specifically, whenever court approval or other legal process is required to compel production of certain forms of data, law enforcement and intelligence agencies should not be permitted to collect or otherwise access that data via contract, license or commercial purchases (unless first obtaining the same legal process, such as a court order, as would be required for compelled production). Absent such a requirement, at a minimum OMB should require law enforcement and intelligence agencies to publish information about their data purchases, including the entities from which they purchase access to data, how much they spend on data purchases, what forms of data are purchased, and how many Americans' personal information is collected via these purchases.

RFI Question 1: *How does AI potentially exacerbate privacy risks associated with agency handling of CAI containing PII?*

a. What are the key privacy risks associated with agencies' handling of CAI containing PII that OMB should consider and why?

Some of the most sensitive and potentially exploitable information about people can be obtained through commercial actors, particularly data brokers. This information can include people's location information (real time and historical), home address, workplace, online or offline activities, and financial information like purchasing habits (which can reveal sensitive attributes such as medical conditions and treatments, or sexual orientation). Consequently, a wide variety of privacy risks are associated with CAI. Further, as repeatedly documented, proliferation of CAI in open markets creates concerns for public safety and national security, as seen from use of such information for stalking to providing ability to track U.S. military personnel.[6] The government should consider these harms in assessing limits to agency handling of CAI containing PII, particularly when that PII is sensitive.

[6] See e.g.: Justin Sherman, *People Search Data Brokers, Stalking, and 'Publicly Available Information' Carve-Outs*, Lawfare (Oct 30, 2023) https://www.lawfaremedia.org/article/people-search-data-brokers-stalking-and-publicly-available-information-carve-outs [https://perma.cc/2573-GPVM]; Dhruv Mehrotra & Dell Cameron, *Anyone Can Buy Data Tracking US Soldiers and Spies to Nuclear Vaults and Brothels in Germany*, Wired (Nov 19, 2024) https://www.wired.com/story/phone-data-us-soldiers-spies-nuclear-germany/ [https://web.archive.org/web/20241124173713/https://www.wired.com/story/phone-data-us-soldiers-spies-nuclear-germany/].

The use of AI exacerbates some of the privacy risks raised by CAI. AI systems may increase the matching of records across datasets, allowing for more inferences about individuals. In addition to situations where the *accurate* matching of records across datasets may cause harms, increases in the use of AI for matching across datasets may also result in more *erroneous* matches, limiting the utility of the data and potentially causing different kinds of harms to people if incorrect data about them is used to limit their access to services or if agencies take actions based on faulty data.

The use of CAI to train AI models may also introduce privacy harms, if, for example, it results in AI that reveals what should be private information. AI can also exacerbate concerns about data leakage (if AI systems regurgitate information they have learned or stored or training data sets are subject to unauthorized access due to data breaches and cyber attacks) and the re-identifiability of de-identified data. It may be difficult to assess and structure data flows to ensure that AI systems do not incorporate sensitive data they have consumed into their outputs, which may be available to people who would not otherwise have access to that information. This limits the already-meager protections sometimes afforded to individuals whose data is incorporated into large CAI datasets.

However, many of the concerns raised by CAI stand regardless of AI. Key among these are quality and equity concerns that agencies must consider and account for when determining whether to use CAI. Datasets like credit information may exclude context that would better reflect the eligibility of certain populations, such as economically disadvantaged communities, for critical opportunities, while other data sets, like arrest records, may be over-representative of populations such as Black communities and other communities of color. The use of these datasets without accounting for the fact that they are deeply influenced by historic inequities and disparate treatment has the potential to perpetuate and exacerbate these disparities. In addition to limiting the utility of this data, these disparities also underline the need to prevent law enforcement and national security actors from using these sorts of data to circumvent Fourth Amendment and other legal protections, as these discrepancies will lead to disproportionate harms to already marginalized and over-policed communities.

Finally, many data broker practices inhibit accountability for the harms flowing from the use of CAI by government agencies, and those practices are harmful regardless of how the government uses such data, raising ethical concerns about government money supporting the industry. The

industry itself relies on a constantly growing, increasingly untraceable network of third parties leading to a broad erosion of privacy and self-determination.[7] While there may be certain benefits to the use of aggregated information in particular circumstances (such as for research and public health efforts during the height of the Covid-19 pandemic), these instances do not justify the risks to individuals and society that stem from use of CAI that contains PII.[8]

RFI Question 3: *What, if any, changes to its current guidance should OMB consider to improve how agencies address and mitigate the privacy risks that may be associated with their handling of CAI containing PII?*

Agencies should perform and publicly release Privacy Impact Assessments (PIAs) for any system, project, or program that uses CAI, and should incorporate CAI risks into AI Impact Assessments (AIAs) for any AI systems that will use or interact with CAI. PIAs should also be updated to address specific risks related to the use of CAI. For instance, some agency PIA templates currently require agencies to disclose whether or not data comes from publicly available or commercial sources, but do not include additional risk considerations specifically tailored to CAI, such as evaluating how, and with what legal authorities, data brokers collected any information purchased or used by an agency.[9] As noted in CDT's comments to OMB regarding PIAs, these documents should be coordinated and standardized across agencies, to allow agencies to benefit from one another's work for shared systems, to ensure that all agencies are assessing for a shared set of risks, and to make it easier for readers, including other agencies, coordination bodies like OMB, the public, and researchers, to analyze impact assessments.[10]

RFI Question 6(c): *Should agencies disclose to individuals when CAI containing PII is used to inform a decision with respect to those individuals (e.g., a determination of their eligibility for or receipt of a Federal benefit)?*

---

[7] Ridhi Shetty, *CDT Comments to CFPB Lay Out Data Broker Harms That Should Be Held Accountable*, Center for Democracy & Technology (Jul 26, 2023) https://cdt.org/insights/cdt-comments-to-cfpb-lay-out-data-broker-harms-that-should-be-held-accountable/ [https://perma.cc/P2JH-EGEC].

[8] Mana Azarmi & Andrew Crawford, *Report: Use of Aggregated Location Information and COVID-19*, Center for Democracy & Technology (May 29, 2020) https://cdt.org/insights/report-use-of-aggregated-location-information-and-covid-19/ [https://perma.cc/QNX5-PGDJ].

[9] See, for instance: U.S. Department of Homeland Security, *Privacy impact assessment template* (accessed Nov 24, 2024) https://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_template.pdf [https://perma.cc/XB7V-GYCG].

[10] Hannah Quay-de-la-Valle, *CDT Submits Comments on the Value of and Potential Improvements to Privacy Impact Assessments (PIAs) From Federal Government Agencies*, Center for Democracy & Technology (Apr 1, 2024) https://cdt.org/insights/cdt-submits-comments-on-the-value-of-and-potential-improvements-to-privacy-impact-assessments-pias-from-federal-government-agencies/ [https://perma.cc/2LHW-G6FW].

i. What steps could agencies take to provide individuals with an opportunity to seek amendment of the CAI before agencies use it to make such decisions?
ii. What other steps could agencies take to verify accuracy, relevance, timeliness, and completeness of the CAI before using it to make decisions about individuals?

RFI Question 7: *Should agencies establish and maintain comprehensive inventories of CAI containing PII that they handle? Why or why not?*

a. If so, should these agency CAI inventories be publicly available? Why or why not?

i. Are there any categories of CAI containing PII that should not be included in a public inventory? If so, what risks support that exclusion?

ii. How would public CAI inventories be useful to stakeholders?

A throughline across these considerations is transparency of agency use of CAI, a core tenet of responsible and ethical data use. Agencies should disclose their use of CAI, both to individuals whose information is used to inform a determination, and also to the broader public.

As part of that effort, agencies should maintain comprehensive inventories of their use of CAI, including the source of that CAI, its cost, the legal authorizations used to obtain the data originally, legal analysis and interpretations used to determine the legality of CAI purchases,[11] how it is used, a data schema, and any steps the agency has taken to ensure the accuracy and completeness of the information. Agencies should have some mechanism for tracking their data sources for correction and to ensure that data no longer needed is destroyed when appropriate to reduce privacy and security risks. As agencies need to maintain these records as a privacy and security best practice, OMB should ensure these inventories are standardized and publicized to make them more valuable and accessible to all stakeholders. OMB should provide detailed guidance about what the inventories should contain and how they should be formatted and made public, and should consider a centralized inventory repository. Creating such inventories aligns with agencies' existing public transparency obligations under the *E-Government Act of 2002* (PIAs) and the *Advancing American AI Act* (AI use case inventories).

Inventories are a valuable transparency tool for research and public evaluation, but they will not ensure that individuals will be aware when their own information is being used to make

[11] For example, legal analysis of to what degree the Supreme Court's *United States v. Carpenter* ruling limits collection of cell site location information and electronic location data generally. See, *United States v. Carpenter*, 585 U.S. 296 (2018), https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf.

determinations that they will be subject to. Agencies should proactively inform individuals about the use of CAI concerning them and disclose the information itself early on in any processes where it is being used, and enable individuals to access and correct this information throughout the decision making process. The specific avenues of disclosure will depend on the information and how it is being used, but approaches like data dashboards, sharing of pre-filled forms or applications, and prompts to review information in other interactions with individuals should be used. All of these avenues should include information about how to amend the information or, ideally, direct links to the amendment processes.

Finally, agencies should ensure they are using up-to-date versions of any information they acquire. This may mean having access to information rather than storing their own copies or checking their version against the source at the time of use. For instance, if an agency is using credit data, and a person has corrected their credit report since the agency acquired that information, the agency would be using incorrect information unless they have a process for ensuring their data is updated.

RFI Question 8: *Should agencies create periodic reports on their handling of CAI containing PII? Why or why not?*

> *a. If so, what information should be included in these reports, and to whom should OMB direct agencies to send these reports?*
> *b. If so, should agencies make these reports publicly available and by what means (e.g., post them on agency privacy program web pages)?*

Agencies should produce these reports. Agencies should report instances where they have retired a data source due to data quality or ethics concerns so that other agencies using the same source can follow suit. Data breaches and security incidents should be tracked as well. Additionally agencies should report what actions they are taking to protect the privacy and integrity of the CAI they collect and use. Finally, agencies should report their expenditures on CAI, and what other avenues for acquiring data they explored before acquiring CAI, and why those avenues were not pursued. Because public agencies' acquisition of CAI often costs taxpayer dollars could at times be acquired instead from partner agencies, such purchases should be done sparingly and only when the potential benefits justify the financial commitment and privacy and ethical risks. In particular, agencies should be required to consider whether

there are ways to acquire the information they need via interagency sharing and report out if such a source exists and why they chose instead to use CAI.

RFI Question 11: *What, if any, means of interagency information sharing should be considered to allow agencies to report problems with CAI containing PII (e.g., recurring concerns with data quality)?*

OMB should create a centralized reporting mechanism for reporting problems with CAI. This would be useful in ensuring that the information about these problems is easily accessible to other agencies; allowing for centralized oversight of uses and issues by OMB, including allowing OMB to identify trends; and, ideally, allowing for public accountability. For the latter use, OMB should establish a public-facing version of the reporting mechanism and incorporate it into public inventories and reports. Each agency should also maintain a single contact point for CAI concerns, to make notification of concerns straightforward for other agencies, OMB, and the public.

Questions about these comments may be directed to CDT's Director of Equity in Civic Technology Elizabeth Laird at elaird@cdt.org.