epic.org | ELECTRONIC PRIVACY INFORMATION CENTER

COMMENTS OF THE ELECTRONIC PRIVACY INFORMATION CENTER

to the

COMMISSION ON EVIDENCE-BASED POLICYMAKING

Request for Public Comment

November 14, 2016

By notice published on September 14, 2016, the Commission on Evidence-Based Policymaking ("CEP") requests public comments on "how to increase the availability and use of government data in support of evidence-building activities related to government programs and policies, while protecting the privacy and confidentiality of such data."[1] Pursuant to this notice, the Electronic Privacy Information Center ("EPIC") submits these comments to (1) make clear that data can be used both for informed policy-making and for profiling, segmentation, and discrimination; (2) urge the Commission to promote privacy-enhancing techniques ("PETs") that minimize or eliminate Personally Identifiable Information; and (3) propose data use schemes that leave the data with the custodial agencies instead of a central repository.

EPIC is a public interest research center in Washington, D.C. EPIC was established in 1994 to focus public attention on emerging privacy and related human rights issues and to

[1] *Request for Comments for the Commission on Evidence-Based Policymaking*, 81 Fed. Reg. 63,166 (Sep. 14, 2016) [hereinafter "Request for Comments].

protect privacy, the First Amendment, and constitutional values. EPIC has a particular interest in safeguarding personal privacy and preventing harmful data practices. For example, EPIC routinely submits comments to federal agencies, urging them to uphold the Privacy Act and protect individual privacy in mass government databases.[2] EPIC is also a leading consumer advocate before the Federal Trade Commission ("FTC"). EPIC has a particular interest in protecting consumer privacy, and has played a leading role in developing the authority of the FTC to address emerging privacy issues and to safeguard the privacy rights of consumers.[3] In 2014, EPIC submitted extensive comments to the White House Office of Science and Technology Policy, warning of the enormous risk to Americans that current "big data" practices present, and recommending the adoption of privacy-enhancing techniques.[4] EPIC also maintains a webpage on practical privacy tools.[5]

[2] *See, e.g.*, EPIC et al., *Comments on the Terrorist Screening Database System of Records, Notice of Privacy Act System of Records and Notice of Proposed rulemaking, Docket Nos. DHS 2011-0060 and DHS 2011-0061* (Aug. 5, 2011), *available at* http://epic.org/privacy/airtravel/Comments_on_DHS-2011-0060_and_0061FINAL.pdf; EPIC, *Comments on Secure Flight*, *Docket Nos. TSA-2007-28972, 2007-28572* (Sept. 24, 2007), *available at* http://epic.org/privacy/airtravel/sf_092407.pdf; EPIC, *Secure Flights Should Remain Grounded Until Security and Privacy Problems are Resolved, Spotlight on Surveillance Series* (August 2007), *available at* http://epic.org/privacy/surveillance/spotlight/0807/default.html; *Passenger Profiling*, EPIC, http://epic.org/privacy/airtravel/profiling.html (last visited Apr. 3, 2014); *Secure Flight*, EPIC, http://epic.org/privacy/airtravel/secureflight.html (last visited Apr. 3, 2014); *Air Travel Privacy*, EPIC, http://epic.org/privacy/airtravel/ (last visited Apr. 3, 2014).

[3] *See, e.g.*, Letter from EPIC Executive Director Marc Rotenberg to FTC Commissioner Christine Varney, EPIC (Dec. 14, 1995) (urging the FTC to investigate the misuse of personal information by the direct marketing industry), http://epic.org/privacy/internet/ftc/ftc_letter.html; DoubleClick, Inc., *FTC* File No. 071-0170 (2000) (Complaint and Request for Injunction, Request for Investigation and for Other Relief), http://epic.org/privacy/internet/ftc/DCLK_complaint.pdf; Microsoft Corporation, FTC File No. 012 3240 (2002) (Complaint and Request for Injunction, Request for Investigation and for Other Relief), http://epic.org/privacy/consumer/MS_complaint.pdf; Choicepoint, Inc., FTC File No. 052-3069 (2004) (Request for Investigation and for Other Relief) , http://epic.org/privacy/choicepoint/fcraltr12.16.04.html.

[4] [cite to comments]

[5] EPIC Online Guide to Practical Privacy Tools, http://epic.org/privacy/tools.html.

1. Data is the basis of research, innovation, economic growth, and informed policy decisions, but data is also the basis for profiling, tracking, segmentation, and discrimination

Although increased use of administrative and survey data has the potential to improve informed policymaking, there are real risks in combining this data and making it more easily available. Data that is improperly protected can be used by the government and in the private sector for profiling, tracking, and discrimination. The potential use of personal information to make automated decisions and segregate individuals based on secret, imprecise and oftentimes impermissible factors presents clear risks to fairness and due process.

A. Government collection and abuse of data

Today, Americans are in more government databases than ever. Government agencies routinely amass personally-identifiable information ("PII") but absolve themselves of any legal duties or responsibilities to safeguard individual privacy. For example, the Federal Bureau of Investigation's Data Warehouse System hoards individual information, including:

> biographical information (such as name, alias, race, sex, date of birth, place of birth, social security number, passport number, driver's license, or other unique identifier, addresses, telephone numbers, physical descriptions, and photographs); biometric information (such as fingerprints); financial information (such as bank account number); location; associates and affiliations; employment and business information; visa and immigration information; travel; and criminal and investigative history, and other data that may assist the FBI in fulfilling its national security and law enforcement responsibilities.[6]

Incredibly, the agency has exempted itself from Privacy Act requirements that the FBI maintain only "accurate, relevant, timely and complete" personal records.[7] The FBI has also exempted itself from Privacy Act requirements permitting individuals to access and amend

[6] Privacy Act of 1974; System of Records, 77 Fed. Reg. 40,630, 40,631 (July 10, 2012), *available at* http://www.gpo.gov/fdsys/pkg/FR-2012-07-10/pdf/2012-16823.pdf.
[7] 28 C.F.R. §16.96 (v).

inaccurate records.[8] Other agencies, like the Department of Homeland Security and the National Security Agency, have exempted databases containing detailed, sensitive personal information from well-established Privacy Act safeguards.[9] EPIC has routinely objected to agencies gathering personally identifiable information while eschewing privacy protections, noting:

> It is inconceivable that the drafters of the Privacy Act would have permitted a federal agency to propose a profiling system on U.S. citizens and be granted broad exemptions from Privacy Act obligations. Consistent and broad application of Privacy Act obligations are the best means of ensuring accuracy and reliability of the data used in a system that profoundly affects millions of individuals as they travel throughout the United States on a daily basis.[10]

The government also uses predictive analytics to the detriment of millions of individuals. For example, the Department of Homeland Security's TSA PreCheck program collects vast amounts of PII including biometric information to perform a "security threat assessment" of "law enforcement, immigration, and intelligence databases, including a fingerprint-based criminal history check conducted through the Federal Bureau of Investigation."[11] The TSA uses automated data processing to determine which individuals will be scrutinized upon traveling throughout the United States.[12] The decisions are completely opaque and lack an effective recourse option. Remarkably, the TSA itself has lost sensitive personal information that it has

---

[8] *Id*.

[9] *See, e.g*., EPIC et al., *Comments on the Department of Defense Privacy Program* (Oct. 21, 2013), *available at* https://epic.org/privacy/nsa/Coal-DoD-Priv-Program-Cmts.pdf; *see also supra* note 3, *Comments Urging the Department of Homeland Security To (A) Suspend the "Automated Targeting System"*.

[10] EPIC, *Comments on TSA PreCheck Application Program System of Records Notice and Notice of Proposed Rulemaking and TSA Secure Flight System of Records Notice*, 5 (Oct. 10, 2013), *available at* http://epic.org/apa/comments/TSA-PreCheck-Comments.pdf.

[11] Privacy Act of 1974: Implementation of Exemptions; Department of Homeland Security Transportation Security Administration, DHS/TSA-021, TSA PreCheck Application Program System of Records, 78 Fed. Reg. at 55,657 (proposed Sept. 11, 2013), *available at* http://www.gpo.gov/fdsys/pkg/FR-2013-09-11/pdf/2013-22069.pdf.

[12] Privacy Act of 1974; Department of Homeland Security Transportation Security Administration--DHS/TSA—019 Secure Flight Records System of Records, 78 Fed. Reg. 55,270, 55,271 (proposed Sept. 10, 2013), *available at* http://www.gpo.gov/fdsys/pkg/FR-2013-09-10/pdf/2013-21980.pdf.

collected from its employees.[13] The TSA lost a portable drive containing the bank account numbers, Social Security numbers, names and birth dates of more than 100,000 people who worked at the TSA over a three-year period.

It is vitally important to ensure that any data clearinghouse minimizes collection, secures the information that is collected, and prevents abuses of collected data through the use of predictive analytics.

B. The 1965 National Data Center Proposal and the Privacy Act of 1974

This Commission's current efforts echo in many ways the goals of the proposed National Data Center in the 1960s. As Rebecca Kraus wrote:

> Computer technology had improved the efficiency and affordability of research with large data sets, and the expansion of government social programs called for more data and research to inform public policy. As a result, in 1965 social scientists recommended that the federal government develop a national data center that would store and make available to researchers the data collected by various statistical agencies.[14]

A 1965 report prepared by the SSRC Committee on the Preservation and Use of Economic Data noted that federal government statistics were highly decentralized and held by agencies that collected the underlying data as a "by-product of the regulatory process."[15] It recommended the creation of a "Federal Data Center" with the authority to obtain data "produced by all federal agencies."[16] The report also recommended the development of an

---

[13] Thomas Frank, *TSA Seeks Hard Drive, Personal Data on 100,000*, USA TODAY, May 5, 2007, available at http://usatoday30.usatoday.com/news/washington/2007-05-04-harddrive-tsa_N.htm?csp=1.
[14] Rebecca S. Kraus, *Statistical Dèjá Vu: The National Data Center Proposal of 1965 and Its Descendants*, 5 J. Privacy & Confidentiality 1, 1 (2013).
[15] RICHARD RUGGLES, RICHARD MILLER, EDWIN KUH, STANLEY LEBERGOTT, GUY ORCUTT & JOSEPH PECHMAN, REPORT OF THE SSRC COMMITTEE ON THE PRESERVATION AND USE OF ECONOMIC DATA (1965).
[16] *Id.* at 1.

organization that could provide a "clearing house and coordination of requests for data made by individual scholars from Federal agencies."[17]

The proposal was met with public outrage.[18] Congress held hearings at which proponents of the national data center appeared to downplay privacy concerns.[19] In 1973, a federal advisory committee released its report on Records, Computers, and the Rights of Citizens.[20] As EPIC President Marc Rotenberg explained in 2000:

> The purpose was benign. It was believed that such a databank would be very useful to social scientists and others, but the implications were severe. People understood that the collection of these permanent profiles, made possible by computerized automation, would pose a threat to the privacy and liberty of American citizens. The proposal for the National Data Center was withdrawn and over time a comprehensive legal framework—the Privacy Act of 1974—was established to safeguards the rights of American citizens. The Privacy Act imposed on all federal agencies essential privacy rights and responsibilities—"Fair Information Practices"—that would limit would federal agencies could do with personal information and gave every American the right to see the information about them that was collected.[21]

The Privacy Act incorporates the Code of Fair Information Practices that the Health, Education, and Welfare Advisory Committee on Automated Data Systems issued in 1973.[22] The Code of Fair Information Practices ("FIPs") sets out five obligations for all organizations that collect personal data:

1. There must be no personal data record-keeping systems whose very existence is secret.

[17] *Id.* at 2.
[18] *Id.* at 13–17; VANCE PACKARD, THE NAKED SOCIETY (Ig Publishing 2014) (1964).
[19] *Invasions of Privacy: Hearings before the Subcomm. on Administrative Practice and Procedure of the Senate Comm. on the Judiciary*, 89th Cong. (1966), https://www.epic.org/privacy/hew1973report/; *see also* Kraus, supra note 14, at 11.
[20] SECRETARY'S ADVISORY COMM. ON AUTOMATED PERS. DATA SYS, RECORDS, COMPUTERS AND THE RIGHTS OF CITIZENS (1973), https://www.epic.org/privacy/hew1973report/.
[21] *Internet Privacy and Profiling: Hearing before the S. Comm. on Commerce, Sci. & Transp.*, 106th Cong. (2000) (statement of Marc Rotenberg, Director, Electronic Privacy Information Center).
[22] *The Code of Fair Information Practices*, EPIC, http://epic.org/privacy/consumer/code_fair_info.html.

2. There must be a way for a person to find out what information about the person is in a record and how it is used.
3. There must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the person's consent.
4. There must be a way for a person to correct or amend a record of identifiable information about the person.
5. Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data.[23]

In passing the Privacy Act of 1974, Congress found that: (1) individual privacy is "directly affected by the collection, maintenance, use, and dissemination of personal information by Federal agencies"; (2) big data in the government sector "greatly magnified the harm to individual privacy"; (3) misuse of government data can threaten "the opportunities for an individual to secure employment, insurance, and credit, and his right to due process"; (4) privacy is a constitutionally-protected "personal and fundamental right"; and (5) "in order to protect the privacy of individuals identified in information systems maintained by Federal agencies, it is necessary and proper for the Congress to regulate the collection, maintenance, use, and dissemination of information by such agencies."[24]

The United States has been slow to update its privacy laws and companies have been reluctant to implement privacy enhancing technologies—neither an appropriate legal framework

[23] U.S. Dep't. of Health, Education and Welfare, Secretary's Advisory Committee on Automated Personal Data Systems, Records, computers, and the Rights of Citizens viii (1973).
[24] Public Law 93-579, 93rd Congress, S.3418, Privacy Act, Section 2 (a) (Dec. 31, 1974).

or technical framework have been implemented to consistently safeguard individual privacy through the FIPs.

The FIPs appear in various privacy laws and frameworks, such as the Organization for Economic Cooperation and Development ("OECD") Privacy Guidelines,[25] the Privacy Act of 1974,[26] and the European Commission's recent Data Protection Regulation.[27] In the United States, the Consumer Privacy Bill of Rights ("CPBR") is a flexible and adaptable instantiation of the FIPs.

The CPBR provides a comprehensive framework that lists seven substantive privacy protections for consumers: Individual Control, Transparency, Respect for Context, Security, Access and Accuracy, Focused Collection, Accountability.[28] This Commission's efforts to make administrative and survey data available for use in evidence-based policymaking while preserving privacy protections should focus on technology that facilitates the implementation of the privacy protections listed in the CPBR.

The reaction to the proposed National Data Center contains several lessons for this Commission. First, privacy must be an integral component of any effort to streamline access to administrative and survey data. Second, the importance of privacy to the project must be clearly communicated to the public. Third, because the idea of a centralized repository is particularly worrisome, any clearinghouse should leave data with the custodial agencies. And finally, a

[25] OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, *available at* http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html.
[26] Privacy Act of 1974, 5 USC § 552a.
[27] Proposal for a Regulation of the European Parliament and the Council on the protection of individuals with regard to the processing of personal data and the free movement of such data (General Data Protection Regulation), E.C. COM (2012) final, (Jan. 25, 2012), available at http://ex.europa/eu/justice/data-protection/document/review2012/com_2012_11_en.pdf
[28] *Id.*

clearinghouse for government data must operate within the protections provided by the Privacy Act.

2. The Commission should encourage the development and use of privacy-enhancing techniques to maximize the benefits and minimize the risks of greater data access

The Commission should focus on Privacy Enhancing Techniques[29] ("PETs") that "minimize or eliminate the collection of personally identifiable information."[30] The Commission can support and further the work of computer scientists that have created various privacy enhancing mechanisms. Techniques that help obtain the advantages of big data while minimizing privacy risks should be encouraged, but these techniques must be robust, scalable, provable, and practical. We discuss some relevant privacy-enhancing techniques below.

A. Data Minimization

The Commission should incorporate data minimization requirements based on those described by the CPBR. The principles that call for federal agencies to "collect only as much personal data as they need to accomplish purposes specified" and "securely dispose of or de-identify personal data once they no longer need it, unless they are under a legal obligation to do otherwise"[31] applies equally to any use or disclosure of agency data. Data minimization protects the confidentiality of consumer data and also serves important data security purposes. Limiting the amount of personal data that agencies collect, retain, and make available also limits the harm that results from possible data breaches.

[29] We use the word "techniques" instead of the more common "technologies" here to reflect the fact that privacy-enhancing methods do not necessarily have to be technological.
[30] Testimony and Statement for the Record of Marc Rotenberg, Executive Director, EPIC, Hearing on Privacy in the Commercial World, Before the Committee on Commerce, Trade, and Consumer Protection (Mar. 1, 2001), http://epic.org/privacy/testimony_0301.html; *See also* Herbert Burkert, *Privacy Enhancing Technologies: Typology, Critique Vision* in PHIL E AGRE AND MARC ROTENBERG, TECHNOLOGY AND PRIVACY: THE NEW LANDSCAPE 125-42 (MIT Press 1998).
[31] White House, CPBR.

Two examples show how evidence-based policymaking can be done without using any personally identifiable information: the U.S. Courts' federal wiretap reports[32] and the National Oceanic and Atmospheric Administration's ("NOAA's") weather data.

The wiretap reports are annual reports to Congress "concerning intercepted wire, oral, or electronic communications" pursuant to federal and state wiretap laws. Federal law requires the Administrative Office of the United States Courts to report the number of federal and state wiretap applications, authorizations, and denials.[33] The reporting requirement provides a common data set that allows researchers, advocates, and government officials to describe the scope of lawful electronic surveillance in the United States. Because the reports are mandated by law, not voluntary or dependent on private sector data sources such as "transparency reports," the reports are regularly reported and stable over time. The methodology for the reports is transparent, the data is provable, and the reports pose no privacy risk because PII is neither collected nor published.

The NOAA uses weather forecasting data, climate data, and satellite imagery extensively. Its reports as used by fishing, shipping, agriculture, and many associated industries. Its data also supports mission-critical functions, emergency services, and local and state governments. None of this data is PII.

B. Anonymization or "De-Identification" of Data

The Commission should ensure that a clearinghouse uses anonymization techniques that adequately de-identify data so that data cannot be combined with other information for re-identification. Because not all de-identification techniques adequately anonymize data, it is

[32] U.S. COURTS, WIRETAP REPORTS, http://www.uscourts.gov/statistics-reports/analysis-reports/wiretap-reports (last visited Nov. 14, 2016).
[33] Omnibus Crime Control and Safe Streets Act of 1968, Pub. L. No. 90-351, § 802, 82 Stat. 197 (codified at 18 U.S.C. § 2519).

important that the process employed is robust, scalable, transparent, and shown to provably prevent the identification of consumer information.[34]

Many companies claim to anonymize or de-identify personal information by aggregating it or assigning pseudonyms to it. Behavioral advertising companies routinely claim that the use of pseudonymous identifiers renders personal information anonymous.[35] Data brokers also rely on the aggregate nature of their marketing data as a defense against criticism of their privacy practices. However, these claims of anonymization are often deceptive. Widely-publicized anonymization failures have shown that even relatively sophisticated techniques have still permitted researchers to identify particular individuals in large data sets.[36]

EPIC favors techniques to de-identify user data,[37] and many scholars are performing valuable research on various de-identification techniques,[38] but greater clarification and standardization is needed. For example, Distinguished Scientist at Microsoft Research Cynthia Dwork has espoused "differential privacy" as a "privacy-preserving analysis."[39] Differential privacy "ensures that the removal or addition of a single database item does not (substantially)

---

[34] *See generally* EPIC, *Re-identification*, http://epic.org/privacy/reidentification/.
[35] *DMA Interest-Based Advertising (IBA) Compliance Alert & Guidelines for Interest-Based Advertising*, Direct Marketing Assoc, http://www.dmaresponsibility.org/privacy/oba.shtml ("Relevant Ads Using Anonymous Data. IBA relies on anonymous, aggregated data to deliver an ad to a computer based on the computer browser's activity, not the activities of a specific individual. Companies use cookies to make this happen.").
[36] *See, e.g.*, Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* http://dataprivacylab.org/projects/identifiability/paper1.pdf; Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1704 (2010) ("Data can be either useful or perfectly anonymous but never both.").
[37] *See generally* Re-identification, ELEC. PRIVACY INFO. CTR., https://epic.org/privacy/reidentification/ (last visited Nov. 19, 2012).
[38] *See, e.g.,* Cynthia Dwork, *Differential Privacy: A Survey of Results*, in THEORY AND APPLICATIONS OF MODELS OF COMPUTATION 1, 3 (Manindra Agrawal et al. eds., 2008); *see also* Latanya Sweeney, *k-anonymity: A Model for Protecting Privacy*, INT'L J. ON UNCERTAINTY, FUZZINESS AND KNOWLEDGE-BASED SYSTEMS, 10(5), 2002; 557- 570.
[39] Cynthia Dwork, *Differential Privacy: A Survey of Results*, 1, 2008, http://www.cs.ucdavis.edu/~franklin/ecs289/2010/dwork_2008.pdf.

affect the outcome of any analysis."[40] Although not an "absolute guarantee of privacy," differential privacy "ensures that only a limited amount of additional risk is incurred by participating in the socially beneficial databases."[41]

Jeff Jonas, Chief Scientist for the IBM Analytics Groups, describes the need to "bake in" privacy protection by, for example, "the ability to anonymize the data at the edge, where it lives in the host system, before you bring it together to share it and combine it with other data."[42] The Commission should focus on improving anonymization techniques to not only increase its effectiveness but also to expand the use cases for anonymization.

3. A clearinghouse should leave the data with the custodial agencies instead of storing data in a central repository

The Commission asks in question 11:

> *How might integration of administrative and survey data in a clearinghouse affect the risk of unintentional or unauthorized access or release of personally-identifiable information, confidential business information, or other identifiable records? How can identifiable information be best protected to ensure the privacy and confidentiality of individual or business data in a clearinghouse?*

EPIC addresses this point to stress that a data clearinghouse should not be a central repository of data. A central database would increase the risk of data breach and insider misuse. It would also be more likely to lead to the kinds of perceptions that led to the demise of the 1965 National Data Center.

The 2015 data breaches at the Office of Personnel Management (OPM), which compromised the personal data of 21.5 million people, including 1.8 million people who did not

[40] *Id*. at 2.
[41] *Id*. at 2-3.
[42] IBM's Jeff Jonas on Baking Data Privacy into Predictive Analytics, *Data Informed,* Nov. 20, 2013, http://data-informed.com/ibms-jeff-jonas-baking-data-privacy-predictive-analytics/#sthash.hBM0lg1N.dpuf

apply for background checks,[43] illustrate the dangers of holding administrative and survey data in a single location. The OPM breach exposed sensitive background investigation data spanning three decades.[44] OPM warns on its website:

> If you underwent a Federal background investigation in 2000 or afterwards (which occurs through the submission of forms SF-86, SF-85, or SF-85P for either a new investigation or a reinvestigation), it is highly likely that you are impacted by the incident involving background investigations. If you underwent a background investigation prior to 2000, you still may be impacted, but it is less likely.[45]

The fingerprints of 5.6 million people were also stolen in the data breach.[46]

Though it may be difficult to imagine, the OPM breach could have been worse if the OPM had held the disparate types of information contemplated in a clearinghouse of administrative and survey data. The more information a database holds, and the more information that resides in the same place, the greater the amount of information that will be disclosed in a breach.

Unauthorized insider access is also a greater threat when data sets are combined into a central location. Criminal dockets contain numerous examples of government employees prying for entertainment or profit. Police officers and deputy sheriffs,[47] customs officers,[48] corrections

---

[43] Dan Goodin, *Call it a "Data Rupture": Hack Hitting OPM Affects 21.5 Million*, ARS TECHNICA (July 9, 2015), http://arstechnica.com/security/2015/07/call-it-a-data-rupture-hack-hitting-opm-affects-21-5-million/.

[44] Andrea Shalal & Matt Spetalnick, *Data Hacked from U.S. Government Dates Back to 1985: U.S. Official*, REUTERS (June 5, 2015), http://www.reuters.com/article/us-cybersecurity-usa-idUSKBN0OL1V320150606.

[45] Office of Personnel Management, *Cybersecurity Resource Center*, https://www.opm.gov/cybersecurity/cybersecurity-incidents/.

[46] Andrea Peterson, *OPM Says 5.6 Million Fingerprints Stolen in Cyberattack, Five Times as Many as Previously Thought*, WASH. POST (Sep. 23 2015), https://www.washingtonpost.com/news/the-switch/wp/2015/09/23/opm-now-says-more-than-five-million-fingerprints-compromised-in-breaches/.

[47] *See, e.g.*, United States v. Black, No. 1:14-cr-00012 (D. Colo. 2014) (running license plates against motor vehicle databases to help a drug-dealing relative determine whether certain vehicles were unmarked police cars); United States v. Cave, No. 8:12-cr-00417 (D. Neb. 2013) (running state Criminal Justice Information Systems (CJIS) searches on behalf of car dealerships seeking to repossess vehicles); United States v. Nowlin, No. 1:12-cr-00513 (D. Md. 2013) (police officer accessing a motor vehicle

officers,[49] Veterans Administration employees,[50] Social Security Administration employees,[51] and many IRS employees[52] have been convicted of access to data for unauthorized purposes. If a clearinghouse of administrative and survey data is created, it is a certainty that someone will look at it despite criminal penalties for doing so. If the data is spread out among the custodial agencies, inaccessible to a single login, the risk of disclosure from insider prying will be minimized.

Finally, the specter of a single database collecting all the government's data about a person is exactly the kind of proposal that led to the demise of the National Data Center and the enactment of the Privacy Act of 1974. Even if data is de-identified—and de-identification would be much more difficult when all data is collected together instead of subsets—many will fear, justifiably so, the uses that such a database might be put to.

4. Conclusion

The use of administrative and survey data has great potential for informed, fact-based policymaking. But it also has the potential to harm privacy and liberty interests. EPIC asks the Commission to encourage the development and use of PETs, including data minimization and robust de-identification of data, in any plan for a data clearinghouse. EPIC also urges the Commission to adopt data use schemes that leave the data with the custodial agencies instead of a central repository.

---

database on behalf of a drug dealer); United States v. Green, No. 4:10-cr-00059 (S.D. Tex. 2010) (sheriff's deputy selling information from the National Crime Information Center (NCIC) database).

[48] *See, e.g.*, United States v. Ben-Shabat, No. 4:09-cr-02180 (D. Ariz. 2010) (customs officer accessing databases to gather information on a company with whom the officer was involved in a legal dispute); United States v. Yanez-Camacho, No. 3:09-cr-02755 (S.D. Cal. 2009).

[49] *See, e.g.*, United States v. Barone, No. 3:08-cr-00174 (D. Conn. 2009).

[50] *See, e.g.*, United States v. Dubree, No. 1:09-cr-00067 (D. Md. 2009) (Veterans Administration employee accessing a co-worker's medical records).

[51] *See, e.g.*, United States v. Wilson, No. 1:09-cr-00662 (D. Md. 2010) (accessing Social Security Administration records for the information necessary to take out a credit card in someone else's name).

[52] *See, e.g.*, United States v. Harris, No. 5:14-cr-00120 (N.D. Tex. 2015); United States v. Krien, No. 2:08-cr-20148 (Kan. 2009); United States v. Supple, No. 3:08-cr-00029 (D. Conn. 2008); United States v. Orr, No. 2:07-cr-00016 (E.D. Ky. 2007); United States v. Jones, No. 1:06-cr-00169 (W.D. Mo. 2006).

Respectfully Submitted,

/s/ Marc Rotenberg
Marc Rotenberg
EPIC President and Executive Director

/s/ James Graves
James Graves
EPIC Law and Technology Fellow