COMMENTS OF THE ELECTRONIC PRIVACY INFORMATION CENTER

to

THE OFFICE OF SCIENCE AND TECHNOLOGY POLICY

Request for Information: Big Data and the Future of Privacy

April 4, 2014

By notice published on March 4, 2014, the Office of Science and Technology Policy ("OSTP") requests public comment on "big data."[1] Pursuant to OSTP's notice, the Electronic Privacy Information Center ("EPIC") submits these comments to: (1) warn the OSTP about the enormous risk to Americans in the current "Big Data" environment; (2) make clear that the challenges of Big Data are not new; (3) call for the swift enactment of the Consumer Privacy Bill of Rights ("CPBR") and the end of opaque algorithmic profiling; (4) highlight the need for stronger privacy safeguards for "Big Data"; and (5) draw attention to international frameworks that provide strong models for safeguarding privacy.

EPIC is a public interest research center in Washington, DC. EPIC was established in 1994 to focus public attention on emerging civil liberties issues and to 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 has adamantly opposed government use of "risk-based" algorithmic profiling.[3] EPIC highlighted the problems inherent in profiling programs like the Department of Homeland Security's ("DHS") Secure Flight in previous testimony and comments. In testimony before the National Commission on Terrorist Attacks Upon the United States (more commonly known as "the 9/11 Commission"), EPIC President Marc Rotenberg explained, "there are specific problems with information technologies for monitoring, tracking, and profiling. The techniques are imprecise, they are subject to abuse, and they are invariably applied to purposes other than those originally intended."[4] 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

---

[1] Government "Big Data," 79 Fed. Reg. 12,251 (Mar. 4, 2014).

[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.*, EPIC et al., *Comments Urging the Department of Homeland Security To (A) Suspend the "Automated Targeting System" As Applied To Individuals, Or In the Alternative, (B) Fully Apply All Privacy Act Safeguards To Any Person Subject To the Automated Targeting System* (Dec. 4, 2006), *available at* http://epic.org/privacy/pdf/ats_comments.pdf; EPIC, *Comments on Automated Targeting System Notice of Privacy Act System of Records and Notice of Proposed Rulemaking, Docket Nos. DHS-2007-0042 and DHS-2007-0043* (Sept. 5, 2007), *available at* http://epic.org/privacy/travel/ats/epic_090507.pdf. *See also*, *Automated Targeting System*, EPIC, https://epic.org/privacy/travel/ats/.

[4] Marc Rotenberg, President, EPIC, *Prepared Testimony and Statement for the Record of a Hearing on Security & Liberty: Protecting Privacy, Preventing Terrorism Before the National Commission on Terrorist Attacks Upon the United States* (Dec. 8, 2003), *available at* http://www.epic.org/privacy/terrorism/911commtest.pdf.

a leading role in developing the authority of the FTC to address emerging privacy issues and to safeguard the privacy rights of consumers.[5]

On January 17, 2014, President Obama announced a plan to take a comprehensive look at the privacy implications of "Big Data."[6] Almost immediately after the White House announcement, EPIC, joined by a coalition of consumer privacy, public interest, scientific, and educational organizations, petitioned OSTP to meaningfully engage the public by accepting public comments on Big Data and the Future of Privacy.[7] The Privacy Coalition urged OSTP to involve the public because it is the public's privacy and future that is at stake when the government and private companies amass big data obtained from the public. The Privacy Coalition encouraged OSTP to consider an array of big data privacy issues, including:

(1)What potential harms arise from big data collection and how are these risks currently addressed?
(2) What are the legal frameworks currently governing big data, and are they adequate?
(3) How could companies and government agencies be more transparent in the use of big data, for example, by publishing algorithms?
(4) What technical measures could promote the benefits of big data while minimizing the privacy risks?
(5) What experience have other countries had trying to address the challenges of big data?
(6) What future trends concerning big data could inform the current debate?

Less than a month after the Coalition filed its petition, the White House announced this public comment opportunity. EPIC appreciates this effort as well as related efforts to encourage public comments on this important policy process.[8]

As discussed below in detail, private organizations and government entities are amassing data with little understanding of the consequences and too few safeguards. In many instances, the organizations gathering the Big Data obtain the benefits, but the individuals bear the consequences.[9] This leads to asymmetries of power and new more subtle means of control. We urge OSTP to incorporate the following observations and recommendations into its final report.

1. The current "Big Data" environment poses enormous risk to Americans

The ongoing collection of personal information in the United States without sufficient privacy safeguards has led to staggering increases in identity theft, security breaches, and financial fraud. Additionally, the 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. Far too many organizations collect detailed personal information and use it with too little regard for the consequences. The current Big Data environment is plagued by data breaches and discriminatory uses of predictive analytics.

---

[5] *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.

[6] John Podesta, *Big Data and the Future of Privacy*, THE WHITE HOUSE BLOG (Jan. 23, 2014, 3:30 PM), http://www.whitehouse.gov/blog/2014/01/23/big-data-and-future-privacy.

[7] EPIC et al., Petition for OSTP to Conduct Public Comment Process on Big Data and the Future of Privacy, Feb. 10, 2014, http://epic.org/privacy/Ltr-to-OSTP-re-Big-Data.pdf.

[8] *Join the Conversation: Big Data, Privacy, and What it Means to You*, THE WHITE HOUSE, http://www.whitehouse.gov/issues/technology/big-data-review (last visited Apr. 3, 2014).

[9] *Big Data and the Future of Privacy*, EPIC, http://epic.org/privacy/big-data/default.html (last visited Apr. 4, 2014).

The use of predictive analytics by the public and private sector undermines our freedom of association. Our online social connections, participation in online debates, and our interests expressed through our online activities can now be used by the government and companies to make determinations about our ability to fly, to obtain a job, a clearance, or a credit card. The use of our associations in predictive analytics to make decisions that have a negative impact on individuals directly inhibits freedom of association. It chills online interaction and participation when those very acts and the associations they reveal could be used to deny an individual a job or flag an individual for additional screening at an airport because of the determination of an opaque algorithm, that may consider a person's race, nationality, or political views.

The ability to predict sensitive data and reveal associations raises the potential for abuse by both the government and the private sector. The information gleaned from predictive analytics could be used in a variety of ways to skirt current legal protections regarding, for example, fairness in housing and employment and First Amendment freedoms of religion and association.[10]

A. Commercial Institutions Collecting Data Have Insufficient Data Security to Protect Americans' Privacy

Over the past year, many disastrous data breaches have occurred. During the busy holiday shopping season, millions of American customers who shopped at Target and Neiman Marcus suffered data breaches. Target suffered a data breach that affected nearly 70 million after its point-of-sale terminals were hacked and compromised because of its own insufficient security standards.[11] This included the account data for roughly 40 million account holders, including their credit and debit card numbers, expiration dates, the three-digit CVV security code, and even PIN data.[12] The customers of Neiman Marcus suffered a very similar data breach in which 1.1 million debit and credit card numbers were compromised.[13]

Last September, a data breach at Adobe exposed the user account information of 38 million users.[14] The breach resulted in the theft of close to 3 million customer credit card numbers.[15] The user account information was similarly exposed in a data breach of LivingSocial that compromised the data of nearly 50 million users.[16] Government agencies routinely lose control of the databases containing detailed personal information they have acquired in the "big data" environment.[17]

---

[10] Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms* 99-101 (Public Law & Legal Theory Research, Working Paper No. 13-64), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2325784.

[11] Target: data breach FAQ, https://corporate.target.com/about/shopping-experience/payment-card-issue-FAQ.

[12] Sarah Perez, *Target's Data Breach Gets Worse: 70 Million Customs Had Info Stolen, Including Names, Emails, and Phones*, TechCrunch, Jan. 10, 2014, http://techcrunch.com/2014/01/10/targets-data-breach-gets-worse-70-million-customers-had-info-stolen-including-names-emails-and-phones/.

[13] Elizabeth A. Harris, Nicole Perlroth & Nathaniel Popper, *Neiman Marcus Data Breach Worse Than First Said*, NYTimes, Jan. 23, 2014, http://www.nytimes.com/2014/01/24/business/neiman-marcus-breach-affected-1-1-million-cards.html.

[14] Brian Krebs, *Adobe Breach Impacted at Least 38 Million Users*, Oct. 29, 2013, Krebs on Security, http://krebsonsecurity.com/2013/10/adobe-breach-impacted-at-least-38-million-users/.

[15] *Id*.

[16] Nicole Perlroth, *LivingSocial Hack Exposes Data for 50 Million Customers*, N.Y. Times, Apr. 26, 2013.

[17] *See, e.g.*, U.S. GOVT' ACCOUNTABILITY OFFICE, GAO-14-487T, INFORMATION SECURITY: FEDERAL AGENCIES NEED TO ENHANCE RESPONSES TO DATA BREACHES (2014), *available at* http://www.gao.gov/assets/670/662227.pdf; William Jackson, *VA Settlement Demonstrates Just How Costly Lax Security Can Be*, GCN, Feb. 2, 2009, http://gcn.com/Articles/2009/02/02/VA-data-breach-suit-settlement.aspx; Majority Staff of H. COMM. ON OVERSIGHT AND GOV'T REFORM, *Information Security Breach at TSA: The Traveler Redress Website* (January 2008), *available at* http://web.archive.org/web/20080131043651/http://oversight.house.gov/documents/20080111092648.pdf; Spencer S. Hsu, *TSA Hard Drive With Employee Data Is Reported Stolen,* WASHINGTON POST (May 5, 2007), http://www.washingtonpost.com/wp-dyn/content/article/2007/05/04/AR2007050402152.html.

In addition to the failure of organizations to adequately safeguard the information they collect, many private companies and government agencies now use opaque and often imprecise techniques that make determinations about individuals that carry real consequences. "Predictive analytics" use algorithms on vast amounts of data to unearth correlations that would otherwise remain hidden.[18] Often, the algorithms leverage seemingly innocuous information to make predictions about sexuality, whether a woman is pregnant, political leanings, and more. One of the more problematic uses of predictive analytics is preemptive predictions that make a specific determination about an individual.

Preemptive predictions limit a person's options by assessing "the likely consequences of allowing or disallowing a person to act in a certain way."[19] Preemptive predictions are made from the perspective "of the state, a corporation, or anyone who wishes to prevent or forestall certain types of action."[20] Examples of preemptive predictions include inclusion on a no-fly list and determinations of credit worthiness. Preemptive predictions are particularly problematic because they are often completely automated decisions made behind a veil of secrecy that lack clear or effective recourse for those individuals who feel they have been wronged by the decision.

The private sector uses big data analytics to make important decisions that affect individuals. A digital lending company has established a loan and credit scoring service that uses big data analytics to assess a person's credit worthiness.[21] The company collects data from social networks, among other sources, to make the automated determination in seconds using a self-learning algorithm.[22]

Even when predictive analytics are not used to make a determination about an individual, they still can be problematic by predicting and, in some instances, revealing sensitive information. The retail chain Target used predictive analytics to predict which female customers were pregnant.[23] This information was given to marketers who revealed the pregnancy of a young woman prior to her telling her parents.[24]

Often, the companies and institutions that are the victims of large-scale data breaches make efforts after-the-fact to improve security and privacy. But this leaves numerous other entities still exposing the personal information of its customers. This problem will only get worse because as John Podesta stated, "There is no question that there is more data than ever before, and no sign that the trajectory is slowing its upward pace."[25]

B. Students are Particularly Vulnerable to Big Data Privacy Risks

Recent large-scale security breaches at educational institutions have compromised student (and faculty) privacy. Last month, a University of Maryland ("UMD") database containing 309,079 student, faculty, staff, and personnel records was breached; the "breached records included name, Social Security number, date of birth, and University identification number" and included records covering a span of 20 years.[26] The university acknowledged that it could have implemented privacy enhancing techniques by purging some of those records "long before the breach."[27] Soon after the UMD breach, Indiana University reported that it had stored names, addresses, and Social

---

[18] VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 11-12 (Houghton Mifflin Harcourt 2013).
[19] Ian Kerr & Jessica Earle, *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy* 66 Stan. L. Rev. Online 65, 67 (2013).
[20] *Id.*
[21] Kreditech: Digital Lending, https://www.kreditech.com/loan-and-credit-scoring/.
[22] *Id.*
[23] Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. Times, Feb. 16, 2012, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.
[24] *Id.*
[25] Counselor John Podesta, Remarks at the White House/MIT "Big Data" Privacy Workshop (Mar. 3, 2014), *available at* http://www.whitehouse.gov/sites/default/files/docs/030414_remarks_john_podesta_big_data.pdf.
[26] Letter from President Loh, Letter from Brian D. Voss concerning UMD Data Breach, http://www.umd.edu/datasecurity/.
[27] Mark Albert, *UMD Testifies to Congress on Massive Data Breach*, WUSA 9, Mar. 27, 2014, http://www.wusa9.com/story/news/local/2014/03/26/university-of-maryland-congress-data-breach/6942023/.

Security numbers for "approximately 146,000 students and recent graduates" in an "insecure location" for almost a year, thus potentially exposing students to identity theft and other forms of fraud.[28] Johns Hopkins University also recently experienced a breach that compromised the names, contact information, and "student-entered comments" of approximately 850 students that were enrolled over a seven-year span.[29] Hackers posted information stolen from the breach, including employee information, on the internet. In response to the breach, Johns Hopkins is exploring privacy enhancing techniques, such as deleting outdated information.[30] These examples illustrate that Big Data places students at risk because schools are not using adequate security standards to protect student records.

Additionally, the mass collection of student information has led to the creation of student dossiers over which students have little to no control. For example, statewide longitudinal databases collect troves of student information comprised of "preschool, K-12, and postsecondary education as well as workforce data."[31] A 2009 Fordham Law School report analyzing statewide longitudinal databases highlights that (1) "most states collected information in excess of what is needed" for government reporting requirements"; (2) student databases "generally had weak privacy protections"; (3) "many states do not have clear access and use rules regarding the longitudinal database"; (4) most states "fail to have data retention policies"; and (5) "several states . . . outsource the data warehouse without any protections for privacy in the vendor contract."[32] Because statewide longitudinal databases collect so much student information and because that information is not adequately protected, Big Data in student statewide longitudinal databases significantly raises the risks that students will be stigmatized throughout their academic career and in the workforce.

Last year, EPIC testified before the Colorado State Board of Education and discussed the growing privacy risks that students face as private companies routinely collect sensitive student records. EPIC discussed how private companies might access extensive disciplinary records, and even facilitate "principal watch lists."[33]

C. Government Collection of Big Data is Particularly Problematic

The government has also abused Big Data. Documents obtained by EPIC through a Freedom of Information Act request show that the Census Bureau provided the Department of Homeland Security statistical data on people who identified themselves on the 2000 census as being of Arab ancestry.[34] The DHS agent who requested the census data explained that it was needed to determine which languages signage should be posted in at major international airports.[35] However, there was no indication that DHS requested similar information about any other ethnic groups.[36] The ultimate abuse of Census information came during World War II, when the Census Bureau provided statistical information to help the War Department round up more than 120,000 innocent Japanese Americans and confine them to internment camps.

Today, Americans are in more government databases than ever. Government agencies routinely amass 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:

---

[28] Indiana University Reports Potential Data Exposure, Feb. 25, 2014, http://news.iu.edu/releases/iu/2014/02/data-exposure-disclosure.shtml.
[29] Johns Hopkins Statement: Breach of a University Server, Mar. 7, 2014, http://releases.jhu.edu/2014/03/07/server-breach/.
[30] *Id*.
[31] *Statewide Longitudinal Data Systems*, EDUCATION DEPARTMENT, https://www2.ed.gov/programs/slds/factsheet.html (last visited Apr. 3, 2014).
[32] CHILDREN'S EDUCATIONAL RECORDS AND PRIVACY: A STUDY OF ELEMENTARY AND SECONDARY SCHOOL STATE REPORTING SYSTEMS, EXECUTIVE SUMMARY (Fordham Law Ctr. on Law and Info. Policy, 2009).
[33] Testimony and Statement for the Record, Khaliah Barnes, EPIC Administrative Law Counsel, Study Session Regarding inBloom, Inc., May 16, 2013, *available at* https://epic.org/privacy/student/EPIC-Stmnt-CO-Study-5-13.pdf.
[34] *Freedom of Information Documents on the Census: Department of Homeland Security Obtained Data on Arab Americans From Census Bureau*, EPIC, http://epic.org/privacy/census/foia/ (last visited Apr. 3, 2014).
[35] EPIC FOIA documents: Email exchange between DHS and Census Bureau, http://epic.org/privacy/census/foia/census_emails.pdf.
[36] *Id.*

> 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.[37]

Incredibly, the agency has exempted itself from Privacy Act requirements that the FBI maintain only "accurate, relevant, timely and complete" personal records.[38] The FBI has also exempted itself from Privacy Act requirements permitting individuals to access and amend inaccurate records.[39] 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.[40] 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.[41]

Like the private sector, 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."[42] The TSA uses automated data processing to determine which individuals will be scrutinized upon traveling throughout the United States.[43] The decisions are completely opaque and lack an effective recourse option. Remarkably, the TSA itself has lost sensitive personal information that it has collected from its employees.[44] 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 update current privacy laws to minimize collection, secure the information that is collected, and prevent abuses of collected data through the use of predictive analytics.

---

[37] 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.
[38] 28 C.F.R. §16.96 (v).
[39] *Id*.
[40] *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"*.
[41] 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.
[42] 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.
[43] 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.
[44] 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.

2. The Challenges that Big Data Present Are Not New

Many of the problems that Americans are confronting today were anticipated when Congress first addressed the challenges of "Big Data" and automating personal information with the Privacy Act of 1974. The Privacy Act incorporates the Code of Fair Information Practices that the Health, Education, Welfare Advisory Committee on Automated Data Systems issued in 1973.[45] The Code of Fair Information Practices 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.
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.[46]

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 big 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."[47]

The findings in the US Privacy Act of 1974 make clear the risks of "big data," long before the term was used.[48] However, the United States has been slow to update its privacy laws. Other countries and regions are moving more effectively to respond to the modern challenge of big data. For example, the European Union Data Protection Directive of 1995 actually anticipated the problem of secretive decisionmaking that would undermine fairness.[49] The right of access, familiar to many in the US, is not limited to simply knowledge about the personal data that is collected but also to how the data is used. According to Article 12 of the Directive:

> *Right of access*
>
> Member States shall guarantee every data subject the right to obtain from the controller:
>
> (a) without constraint at reasonable intervals and without excessive delay or expense: confirmation as to whether or not data relating to him are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed; communication to him in an intelligible form of the data undergoing processing and of any available information as to their source; *knowledge of the logic*

[45] *The Code of Fair Information Practices*, EPIC, http://epic.org/privacy/consumer/code_fair_info.html.
[46] 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).
[47] Public Law 93-579, 93rd Congress, S.3418, Privacy Act, Section 2 (a) (Dec. 31, 1974).
[48] In the 1960s and 1970s, commentators and policy makers were more likely to say "databanks" or "databases." *See, e.g.*, ARTHUR R. MILLER, THE ASSAULT ON PRIVACY: COMPUTERS, DATA BANKS, AND DOSSIERS (University of Michigan Press 1971); ALAN F. WESTIN, PRIVACY AND FREEDOM (Bodley Head 1970).
[49] S*ee generally EU Data Protection Directive,* EPIC, http://epic.org/privacy/intl/eu_data_protection_directive.html.

> *involved in any automatic processing of data* concerning him at least in the case of the automated decisions referred to in Article 15(1);[50]

As a document prepared for Europeans explains:

> You must also have access to the logic on which automated decisions are based. Decisions, which significantly affect the data subject, such as the decision to grant a loan or issue insurance, might be taken on the sole basis of automated data processing. Therefore, the data controller must adopt suitable safeguards, such as giving the data subject the opportunity to discuss the rationale behind the data collected or to contest decisions based on inaccurate data.[51]

Although the Privacy Act of 1974 anticipated many of the challenges that Big Data present, the current legal frameworks fail to safeguard individual privacy by adequately implementing Fair Information Practices ("FIPs") and adhering to privacy enhancing techniques. Because Big Data has threatened individual privacy for many years, and the risks to Americans increase daily, it is imperative that this Administration confronts Big Data problems expeditiously. Among the changes that are needed, the law should be updated to guarantee algorithmic transparency.

3. Congress Should Swiftly Enact the Consumer Privacy Bill of Rights and the Government Should Immediately Suspend its "Risk Based" Profiling Programs

In 2012, President Obama announced the Consumer Privacy Bill of Rights ("CPBR").[52] It is a critical policy framework that provides a blueprint for protecting privacy in the modern age. Based on FIPs, the CPBR is a framework that grants consumer rights and places obligations on private companies collecting consumer information:

- Individual Control: Consumers have a right to exercise control over what personal data companies collect from them and how they use it.
- Transparency: Consumers have a right to easily understandable and accessible information about privacy and security practices.
- Respect for Context: Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data.
- Security: Consumers have a right to secure and responsible handling of personal data.
- Access and Accuracy: Consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity of the data and the risk of adverse consequences to consumers if the data is inaccurate.

---

[50] EU Directive 95/46/EC—The Data Protection Directive, art 15 (1), 1995 (emphasis added), http://www.dataprotection.ie/docs/EU-Directive-95-46-EC--Chapter-2/93.htm. Article 15(1) is expansive and includes "data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."

[51] EUROPA, Data Protection in the European Union, 9, *available at* http://ec.europa.eu/justice/policies/privacy/docs/guide/guide-ukingdom_en.pdf.

[52] White House, Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Economy, Feb. 23, 2012, http://www.whitehouse.gov/sites/default/files/privacy-final.pdf [hereinafter White House, CPBR]; *see also White House Sets Out Consumer Privacy Bill of Rights*, EPIC, http://epic.org/2012/02/white-house-sets-out-consumer-.html (last visited Apr. 4, 2014).

- Focused Collection: Consumers have a right to reasonable limits on the personal data that companies collect and retain.

- Accountability: Consumers have a right to have personal data handled by companies with appropriate measures in place to assure they adhere to the Consumer Privacy Bill of Rights.[53]

The Consumer Data Privacy Report identifies several high-profile privacy challenges, including online advertising, data brokers, and children's privacy. The report encourages online advertising companies to "refrain from collecting, using, or disclosing personal data that may be used to make decisions regarding employment, credit, and insurance eligibility" and cited a "Do Not Track" mechanism as an example of a beneficial privacy-enhancing technology.[54] The report calls on data brokers to "seek innovative ways to provide consumers with effective Individual Control."[55] Finally, the report notes, "the practices in the Consumer Privacy Bill of Rights may require greater protections for personal data obtained from children and teenagers than for adults."[56]

More than two years have passed since President Obama announced the CPBR. But with no action on the recommendations or the framework, the problems with Big Data have increased. Last month, forty consumer privacy organizations urged the White House to work with Congress to propose legislation enacting the Consumer Privacy Bill of Rights into law. The groups stated:

> Never has the need to update the privacy laws of the United States been more urgent. . . . The Consumer Privacy Bill of Rights is a sensible framework that would help establish fairness and accountability for the collection and use of personal information. . . . the key to progress is the enactment by Congress of this important privacy framework. Only enforceable privacy protections create meaningful safeguards.[57]

The time to act is now. The White House must work with Congress to enact the CPBR and protect the privacy of Americans.

Additionally, to combat the problems with preemptive predictions based on government Big Data, the government must immediately cease "risk based" automated profiling. For example, the Department of Homeland Security ("DHS") uses its Automated Targeting System ("ATS") to assign risks to individuals traveling to, from, and throughout the United States.[58] DHS uses PII to determine whether an individual, based on personal immutable characteristics—not conduct—should undergo investigation, monitoring, and denial of her constitutional right to travel.[59] This is almost certainly constitutionally impermissible.[60] Moreover, because ATS risk assessment compares PII of individuals that have no criminal history against "patterns of suspicious activity," this increases the likelihood that CBP and ATS profile innocent individuals of certain racial, ethnic, or religious groups.[61]

---

[53] *See supra* note 52, White House, CPBR. at 1.
[54] *Id.* at 12.
[55] *Id.* at 13.
[56] *Id.* at 15.
[57] Privacy Coalition Letter on Consumer Privacy Bill of Rights, Feb. 24, 2014, *available at* http://epic.org/privacy/Obama-CPBR.pdf.
[58] Notice of Privacy Act System of Records, 77 Fed. Reg. 30297 (proposed May 22, 2012).
[59] *Sáenz v. Roe*, 526 U.S. 489 (1999).
[60] *See supra* note 3, *Comments Urging the Department of Homeland Security To (A) Suspend the "Automated Targeting System." See also* EPIC, *Comments to CBP regarding Automated Targeting System*, June 21, 2012, *available at* http://epic.org/privacy/travel/ats/EPIC-ATS-Comments-2012.pdf.
[61] Dep't of Homeland Sec., U.S. Customs and Border Protection, Privacy Impact Assessment for the Automated Targeting System, DHS/CBP/PIA-006(b), 19 (June 1, 2012), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_cbp_ats006b.pdf.

EPIC and other privacy and civil liberties organizations have repeatedly called for suspension of automated "risk based" profiling.[62] In conducting its review on Big Data in the government sector, the White House should advocate for the immediate suspension of automated "risk based" profiling.

Big companies, including internet advertisers, should also be far more transparent about their profiling and pricing practices. Every indication points to the increasing use of secretive profiles, filled with sensitive data, to make determinations about American consumers.[63] Those practices should end. Companies should not be allowed to make decisions about individuals without setting out in detail the basis for the decision, including the factors that it considered. And government agencies must be on the lookout for the use of factors, such as race, gender, and nationality that are Constitutionally impermissible.

4. Stronger Big Data Privacy Safeguards Are Needed to Protect Individuals

A. *Current Practices*

It is imperative that stronger safeguards are implemented to protect the privacy and personal information of Americans. Transparency is often foregone in order to avoid accountability for the accuracy of the data or for how the data is used. Various entities access Big Data with little accountability. For example, private information aggregators increasingly sell consumer profiles that are not clearly protected under current legal frameworks.

When legal frameworks are inapplicable to new uses of Big Data, companies collecting Big Data are not held accountable to regulatory bodies. Spokeo, a people-finder service, is one such company. Spokeo sells detailed consumer profiles, including emails, physical addresses, phone numbers, marital status, occupation, family background, and more.[64] Although Spokeo profits from selling consumer profiles like credit reporting agencies do, it makes no warrants regarding accuracy of its profiles.[65]

Professor Anita Ramasastry notes that Spokeo and its ilk compile "robust data sets . . . that creditors want, and at present, it is unclear how actively they stop such companies from using this information." Because Spokeo's data sets are "used for a major life decision, such as whether someone might be hired or not, the person affected has no recourse, or ability to correct the errors." Professor Ramasastry has stated that Spokeo and other information aggregators "need[]to be subject to some regulatory scrutiny. At a minimum, consumers should have the ability to see their data, to correct it if needed, and to understand who might be buying their data for commercial purposes." Transparency is not sufficient and mechanisms for oversight are also needed.

Many state education departments, for example, lack adequate oversight of schools' use of technology and outsourcing of student records. Sheila Kaplan, a student privacy advocate and founder of Education New York, has endorsed state education chief privacy officers as an independent mechanism to "oversee, audit, consult, and report on matters that affect privacy and security of school records that contain personally identifiable information."[66]

---

[62] *See, e.g.*, EPIC et al., Letter to DHS Secretary Janet Napolitano, Re: TSA Racial Profiling Audit, Dec. 1, 2011, http://epic.org/privacy/airtravel/12-01-11-Coalition-Racial-Profiling-Audit-DHS-Letter.pdf.

[63] *See* Pam Dixon & Robert Gellman, World Privacy Forum, *The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future* (Apr. 2, 2014), *available at* http://www.worldprivacyforum.org/wp content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf.

[64] Spokeo, www.Spokeo.com (last visited Apr. 4, 2014).

[65] Anita Ramasastry, *The Spokeo Lawsuit and the Perils of the New People Finder Companies*, http://verdict.justia.com/2014/02/11/spokeo-lawsuit-perils-new-people-finder-companies.

[66] MODEL STATE LAW: CHIEF PRIVACY OFFICE FOR EDUCATION ACT § 1 (Sheila Kaplan, Educ. N.Y.), *available at* http://educationnewyork.com/files/CPOforED-2-01.pdf.

B. Support for Privacy Coalition Principles

As a starting point for a policy framework that could address the challenges of Big Data, EPIC supports the Principles set out by the Privacy Coalition and recommends that the White House incorporate these principles in its report:

> **TRANSPARENCY**: Entities that collect personal information should be transparent about what information they collect, how they collect it, who will have access to it, and how it is intended to be used. Furthermore, the algorithms employed in big data should be made available to the public.
>
> **OVERSIGHT**: Independent mechanisms should be put in place to assure the integrity of the data and the algorithms that analyze the data. These mechanisms should help ensure the accuracy and the fairness of the decision-making.
>
> **ACCOUNTABILITY**: Entities that improperly use data or algorithms for profiling or discrimination should be held accountable. Individuals should have clear recourse to remedies to address unfair decisions about them using their data. They should be able to easily access and correct inaccurate information collected about them.
>
> **ROBUST PRIVACY TECHNIQUES**: 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. And solutions that may be many years into the future provide no practical benefit today.
>
> **MEANINGFUL EVALUATION**: Entities that use big data should evaluate its usefulness on an ongoing basis and refrain from collecting and retaining data that is not necessary for its intended purpose. We have learned that the massive metadata program created by the NSA has played virtually no role in any significant terrorism investigation. We suspect this is true also for many other "big data" programs.
>
> **CONTROL**: Individuals should be able to exercise control over the data they create or is associated with them, and decide whether the data should be collected and how it should be used if collected.[67]

These requirements are needed as entities try to take advantage of big data more and more without taking on the responsibility that should come with it. Greater transparency is an important place to start. As a recent Senate Majority report noted about Data Brokers, "Since data brokers generally collect information without the consumers' knowledge, consumers have limited means of knowing how the companies obtain their information, whether it's accurate, and for what purposes they are using it."[68] And public availability of data should not excuse companies or the government from being responsible data stewards. As danah boyd and Kate Crawford note, "The process of evaluating the research ethics cannot be ignored simply because the data is seemingly accessible. Researchers must keep asking themselves – and their colleagues – about the ethics of their data collection, analysis, and publication."[69] This same sentiment should apply to all entities that collect data.

---

[67] Privacy Coalition Letter on Big Data and the Future of Privacy, Mar. 31, 2014, *available at* http://privacycoalition.org/Big.Data.Coalition.Ltr.pdf .

[68] Office of the Oversight and Investigations Majority Staff, "A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes," 5 (Dec. 18, 2013), *available at* http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0d2b3642-6221-4888-a631-08f2f255b577.

[69] danah boyd and Kate Crawford. *Six Provocations for Big Data* at 11. Research paper presented at Oxford Internet Institute's "A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society" (Sep. 21, 2011), http://ssrn.com/abstract=1926431.

C. *Privacy Enhancing Techniques and Other Practices*

There are other recommendations that should be incorporated in the White House report. In the consumer space, University of Washington Law Professor Ryan Calo has suggested the creation of Consumer Subject Review Boards ("CSRBs"). CSRBs would be internal committees within private companies and they would operate under predetermined ethical rules to adequately vet and oversee the big data privacy implications of consumer behavioral research.[70] Among other benefits, CSRBs "could increase regulatory certainty, perhaps forming the basis for an FTC safe harbor if sufficiently robust and transparent" and they "could add a measure of legitimacy to the study of consumers for profit."[71]

Additionally, the public and private sector should implement Privacy Enhancing Technologies ("PETs") that "minimize or eliminate the collection of personally identifiable information."[72] Computer scientists have created various privacy enhancing mechanisms that should be deployed in the Big Data space. Distinguished Scientist at Microsoft Research Cynthia Dwork has espoused "differential privacy" as a "privacy-preserving analysis."[73] Differential privacy "ensures that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis."[74] 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."[75] Current FTC Chief Technologist Latanya Sweeney has created various algorithms that maintain confidentiality by "providing the most general version of the data."[76]

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."[77]

The techniques are particularly important to address the potential abuses of predictive analytics. Where decisions are being made about individuals using predictive analytics, a process is needed to ensure the fairness of the decision. The "technological due process" as described by Danielle Citron provides a good basis for ensuring fairness in automated decisions.[78] Citron suggests audit trails that provide the information used to make a determination would provide transparency to users and a means to affectively challenge these decisions. The audit trails would also provides a means of oversight and accountability in the use of predictive analytics.

5. International Guidelines Provide Models for Better Privacy Protection

In addition to the Organisation for Economic Cooperation and Development Guidelines, there are other international frameworks that can serve as models to protect privacy in big data. In 2012, the European Commission proposed the "EU General Data Protection Regulation," ("GDPR") which has gained support from numerous U.S. consumer organizations. U.S. groups support the Regulation because it "establishes single, national data protection authorities in each [EU] member state," "adopts several innovative approaches to privacy

[70] *Id*.
[71] *Id*.
[72] 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).
[73] Cynthia Dwork, *Differential Privacy: A Survey of Results*, 1, 2008, http://www.cs.ucdavis.edu/~franklin/ecs289/2010/dwork_2008.pdf.
[74] *Id*. at 2.
[75] *Id*. at 2-3.
[76] Latanya Sweeney, *Datafly: a System for Providing Anonymity in Medical Data*, 15, http://dataprivacylab.org/datafly/paper2.pdf.
[77] 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
[78] Danielle Keats Citron, Technological Due Process (2008).

protection, such as privacy by design and privacy by default," and "builds on the right to data deletion."[79] In 2013, the European Parliament Committee approved the regulation. [80]

The Madrid Declaration is an international "commitment to privacy protection" that "reaffirms international instruments for privacy protection, identifies new challenges, and call[s] for concrete actions."[81] Formally endorsed by hundreds of domestic and international civil society groups, privacy experts, and individuals, the Declaration promotes ten propositions concerning data protection.[82] For example, it reaffirms support for Fair Information Practice global implementation, genuine Privacy Enhancing techniques and Privacy Impact Assessments, and "independent data protection authorities."[83] It calls for a moratorium on mass surveillance technology; including body scanners, facial recognition, and RFID tracking, "subject to a full and transparent evaluation by independent authorities and democratic debate."[84]

The Council of Europe Convention 108 is an agreement, signed by the member states of the Council of Europe in 1995, which "protects the individual against abuses which may accompany the collection and processing of personal data and which seeks to regulate at the same time the transfrontier flow of personal data."[85] Convention 108 imposes certain rules regarding the methods by which signatory countries must regulate personal data collection and retention, and also forbids processing of "sensitive" data on a person's race, politics, health, religion, sexual life, criminal record, etc., in the absence of proper legal safeguards. The Convention also "enshrines the individual's right to know that information is stored on him or her and, if necessary, to have it corrected."[86] The Convention still remains the only binding international legal instrument with a worldwide scope of application in the field of data privacy, open to any country, including countries that are not Members of the Council of Europe.[87]

While Convention 108 itself originated in the Council of Europe, the United States' demonstrated interest in the privacy principles protected by the Convention align perfectly. The principles underlying Convention 108 are directly based on the Universal Declaration of Human Rights, adopted by the United Nations in 1948.[88] It was the United States and Eleanor Roosevelt that helped craft the Universal Declaration, and it was the United States that ratified the Council of Europe Convention on Cybercrime and urged its allies to do the same.

Moreover, technical experts and legal scholars have expressed support for the U.S. ratification of Convention 108. On January 28, 2010, twenty-nine members of the EPIC Advisory Board wrote to then Secretary of State Hillary Rodham Clinton to urge that the United States begin the process of ratification of Council of Europe Convention 108.[89] In that letter, the members of the EPIC Advisory Board explained, "Just as communications networks can be used for good and ill, so too can computer technology. It can help sustain aid programs, spur innovation, and encourage economic growth. Or it can track the activities of dissidents, monitor the

---

[79] Letter from U.S. Consumer Organizations on EU General Data Protection Regulation to Jan Philipp Albrecht, Rapporteur, Comm. on Civil Liberties, Justice and Home Affairs, and Lara Comi, Rapporteur, Comm. on Internal Market and Consumer Protection, European Parliament (Sept. 5, 2012), *available at* https://epic.org/privacy/intl/US-Cons-Grps-Support-EU-Priv-Law.pdf.

[80] Press Release, European Parliament, Civil Liberties MEPs Pave the Way for Stronger Data Protection in the EU (Oct. 21, 2013), http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-%2f%2fEP%2f%2fTEXT%2bIM-PRESS%2b20131021IPR22706%2b0%2bDOC%2bXML%2bV0%2f%2fEN&language=EN.

[81] Madrid Privacy Declaration: Global Privacy Standards for a Global World, The Public Voice (Nov. 3, 2009), *available at* http://thepublicvoice.org/madrid-declaration/.

[82] *Id*.

[83] *Id*.

[84] *Id*.

[85] Council of Europe: Treaty Office: Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (last accessed May 9, 2013), *available at* http://conventions.coe.int/Treaty/en/Summaries/Html/108.htm.

[86] *Id*.

[87] *Council of Europe Privacy Convention,* EPIC, http://epic.org/privacy/intl/coeconvention/ (last visited Apr. 4, 2014).

[88] Letter from EPIC Advisory Board to Secretary Clinton, January 28, 2010, *available at* http://epic.org/privacy/intl/EPIC_Clinton_ltr_1-10.pdf.

[89] *Id.*

private lives of citizens, and maintain elaborate systems of identification for laborers and immigrants... the protection of privacy is a fundamental human right. In the 21st century, it may become one of the most critical human rights of all. Civil society organizations from around the world have recently asked that countries which have not yet ratified the Council of Europe Convention 108 and the Protocol of 2001 to do so as expeditiously as possible."[90] The next day, the U.S. Privacy Coalition, comprised of twelve privacy groups, including EPIC, also signed a resolution to the U.S. Senate endorsing Convention 108.[91]

The White House should look to these international models in developing necessary safeguards for the challenge of "Big Data." Technology has outpaced the law, but it is not too late to establish the safeguards that allow for the insights offered by big data, while protecting the fundamental rights of Americans

Conclusion

EPIC appreciates this opportunity to comment and looks forward to continued public engagement on the issue of big data and privacy.

Respectfully submitted,

Marc Rotenberg
EPIC President and Executive Director

Julia Horwitz
EPIC Consumer Protection Counsel

Jeramie Scott
EPIC National Security Counsel

Khaliah Barnes
EPIC Administrative Law Counsel

Electronic Privacy Information Center (EPIC)
1718 Connecticut Avenue, NW, Suite 200
Washington, DC 20009
(202) 483-1140

[90] *Id.*
[91] Privacy Coalition, Resolution, United States Senate, Jan. 29, 2009 *available at* http://privacycoalition.org/resolution-privacy_day.pdf.