COMMENTS OF THE ELECTRONIC PRIVACY INFORMATION CENTER

to

THE NATIONAL COORDNIATION OFFICE (NCO) FOR NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT (NITRD)

Request for Information: National Privacy Research Strategy

October 17, 2014

---

By notice published on September 18, 2014, the National Coordination Office ("NCO") for Networking and Information Technology Research and Development ("NITRD") announced a request for information ("RFI") on a National Privacy Research Strategy. Specifically, NITRD solicits input "on defining the most important goals for privacy in the digital world. As a strategy, the NPRS must focus research activities toward relevant and impactful objectives, and this RFI seeks to inform our understanding of societal needs where privacy-enhancing technologies would be beneficial."[1]

Pursuant to NITRD's notice, the Electronic Privacy Information Center ("EPIC") submits these comments to: (1) highlight current privacy issues; (2) recommend a privacy architecture based on Fair Information Practices and the Consumer Privacy Bill of Rights; and (3) propose specific privacy-enhancing technologies that "minimize or eliminate the collection of personally identifiable information."[2]

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, especially with respect to digital privacy. Earlier this year, 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.[3] EPIC also maintains a webpage on practical privacy tools.[4]

---

[1] Request for Information (RFI)–National Privacy Research Strategy, 79 Fed. Reg. 56,091, 56,092 (Sept. 18, 2014).
[2] At the time of the EPIC filing the President issued a new Executive Order to strengthen consumer privacy. The White House also announced a series of measures to safeguard consumer financial security, including more secure payment systems, efforts to reduce identity theft and support for "algorithmic transparency." EPIC strongly supports this new initiative and urges the NITRD to take into account these new efforts to improve the security of online payment techniques.
[3] EPIC, *Comments on Big Data and the Future of Privacy*, FR Doc. 2014-04660 (Apr. 4, 2014), *available at* http://epic.org/privacy/big-data/EPIC-OSTP-Big-Data.pdf; *See also*, EPIC: Big Data and the Future of Privacy, *available at* https://www.epic.org/privacy/big-data/.
[4] EPIC: Online Guide to Practical Privacy Tools, http://epic.org/privacy/tools.html.

In the RFI, NITRD seeks public input on "privacy objectives," "assessment capabilities," a "multi-disciplinary approach" to "both strengthen privacy and support innovation," and "privacy architectures."[5]

**NITRD Privacy Objectives Should Confront Current Privacy Risks**

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. Personal information is often collected surreptitiously or the collection is obfuscated and used for purposes never intended by the user or consumer. 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 lack of a privacy-protecting architecture and use of privacy-enhancing techniques has created an environment plagued by overcollection, unintended secondary uses of personal information, data breaches, and discriminatory uses of predictive analytics.

*Government and 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 last year, 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.[6] 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.[7] The customers of Neiman Marcus suffered a very similar data breach in which 1.1 million debit and credit card numbers were compromised.[8]

Last September, a data breach at Adobe exposed the user account information of 38 million users.[9] The breach resulted in the theft of close to 3 million customer credit card numbers.[10] The user account information was similarly exposed in a data breach of LivingSocial that compromised the data of nearly 50 million users.[11] Government agencies routinely lose

[5] Request for Information (RFI)–National Privacy Research Strategy, 79 Fed. Reg. 56,091-56,092 (Sept. 18, 2014).
[6] Target: data breach FAQ, https://corporate.target.com/about/shopping-experience/payment-card-issue-FAQ.
[7] 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/.
[8] 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.
[9] 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/.
[10] *Id*.
[11] Nicole Perlroth, *LivingSocial Hack Exposes Data for 50 Million Customers*, N.Y. Times, Apr. 26, 2013.

control of the databases containing detailed personal information they have acquired in the "big data" environment.[12]

Often, the information collected about users and consumers is done surreptitiously. Detailed personal information, including web browsing history, location, and associations are commonly collected by companies without individuals' knowledge. Moreover, this information is regularly utilized for secondary uses beyond the original purpose of the collection.

*The Use of Predictive Analytics Threatens Privacy*

In addition to the lack of transparency surrounding data collection and 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" uses algorithms on vast amounts of data to unearth correlations that would otherwise remain hidden.[13] Often, the algorithms leverage seemingly innocuous information to make predictions about sexuality, whether a woman is pregnant, political leanings, and more.

The use of predictive analytics by the public and private sector undermines freedom of association. Online social connections, participation in online debates, and interests expressed through online activities can now be used by the government and companies to make determinations about individuals' ability to travel, to obtain a job, a clearance, or a credit card. The use of individual 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.

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."[14] Preemptive predictions are made from the perspective "of the state, a corporation,

---

[12] *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.

[13] VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 11-12 (Houghton Mifflin Harcourt 2013).

[14] Ian Kerr & Jessica Earle, *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy* 66 Stan. L. Rev. Online 65, 67 (2013).

or anyone who wishes to prevent or forestall certain types of action."[15] 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.[16] The company collects data from social networks, among other sources, to make the automated determination in seconds using a self-learning algorithm.[17]

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.[18] This information was given to marketers who revealed the pregnancy of a young woman prior to her telling her parents.[19]

The problems created by subpar data security, surreptitious collection, and predictive analytics will only get worse because, as John Podesta, who led the White House's review of big data and privacy, stated, "There is no question that there is more data than ever before, and no sign that the trajectory is slowing its upward pace."[20]

**NITRD Should Implement a Privacy Preserving Architecture Based on Fair Information Practices**

The current data collection architecture is not tenable. Companies seek every opportunity to exploit technology to collect information, while largely abdicating any responsibility for the security of that information or for how that information is used.[21] The companies and institutions that collect data need to take on more accountability and implement a privacy preserving architecture that improves data security and applies privacy enhancing techniques. The United States already has a framework to facilitate the implementation of a privacy preserving architecture—the Fair Information Practices (the "FIPs"). More specifically, the Consumer Privacy Bill of Rights ("CPBR"), which is based on the FIPs and provides a baseline set of privacy protections for users and consumers.

---

[15] *Id.*
[16] Kreditech: Digital Lending, https://www.kreditech.com/loan-and-credit-scoring/.
[17] *Id.*
[18] Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. Times, Feb. 16, 2012, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.
[19] *Id.*
[20] 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.
[21] *See, e.g.*, Alyson Shontell, *Snapchat: If Your Nude Snapchat Photos Get Leaked, It's Not Our Fault*, Business Insider, Oct. 10, 2014, http://www.businessinsider.com/snapchat-if-your-nude-photos-get-leaked-its-not-our-fault-2014-10.

*The Code of Fair Information Practices ("FIPs")*

Congress first addressed the challenges that arise from the collection and automating of 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.[22] 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.[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 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."[24]

The findings in the US Privacy Act of 1974 make clear the risks of collection long before the term "big data" was used.[25] However, 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 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,[26] the Privacy Act of

[22] EPIC: The Code of Fair Information Practices, http://epic.org/privacy/consumer/code_fair_info.html.
[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).
[25] 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).
[26] 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.

1974,[27] and the European Commission's recent Data Protection Regulation.[28] In the US, the CPBR represents the most recent example of 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, and Accountability.[29] The Privacy Research Agenda should focus on technology that facilitates the implementation of the privacy protections listed in the CPBR.

**Privacy Enhancing Techniques and Other Practices**

The NITRD should focus on Privacy Enhancing Technologies ("PETs") that "minimize or eliminate the collection of personally identifiable information."[30] NITRD can support and further the work of Computer scientists that have created various privacy enhancing mechanisms. Distinguished Scientist at Microsoft Research Cynthia Dwork has espoused "differential privacy" as a "privacy-preserving analysis."[31] Differential privacy "ensures that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis."[32] 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."[33] Current FTC Chief Technologist Latanya Sweeney has created various algorithms that maintain confidentiality by "providing the most general version of the data."[34]

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."[35] The NITRD should focus on improving anonymization techniques to not only increase its effectiveness but also to expand the use cases for anonymization.

Techniques are particularly important to address the potential abuses of predictive analytics. Where decisions are being made about individuals using predictive analytics, a process

---

[27] Privacy Act of 1974, 5 USC § 552a.
[28] 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
[29] *Id.*
[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] Cynthia Dwork, *Differential Privacy: A Survey of Results*, 1, 2008, http://www.cs.ucdavis.edu/~franklin/ecs289/2010/dwork_2008.pdf.
[32] *Id*. at 2.
[33] *Id*. at 2-3.
[34] Latanya Sweeney, *Datafly: a System for Providing Anonymity in Medical Data*, 15, http://dataprivacylab.org/datafly/paper2.pdf.
[35] 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

is needed to ensure the fairness and transparency of the decision. The "technological due process" as described by Danielle Citron provides a good basis for ensuring fairness in automated decisions.[36] 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. The NITRD can work to improve the techniques for tracking the use of information in order to facilitate the creation of audit trails. Similarly, the NITRD should focus on creating a set of standards for reviewing predictive algorithms to ensure fairness and prevent profiling and discrimination. Additionally, the NITRD can lead the way in demonstrating how to properly implement algorithmic transparency to provide consumers and users insight into the working of algorithms that make decisions and draws inferences about them.

**Conclusion**

EPIC appreciates the opportunity to comment and looks forward to continued public engagement on the NITRD's National Privacy Research Agenda.

Respectfully Submitted,

Marc Rotenberg
EPIC President and Executive Director

Julia Horwitz
EPIC Consumer Protection Counsel

Khaliah Barnes
EPIC Administrative Law Counsel

Jeramie Scott
EPIC National Security Counsel

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

[36] Danielle Keats Citron, Technological Due Process (2008).