WASHINGTON
LEGISLATIVE OFFICE

March 31, 2014

Office of Science and Technology Policy
Eisenhower Executive Office Building
1650 Pennsylvania Ave. NW.
Washington, DC 20502
bigdata@ostp.gov

RE: Big Data RFI – ACLU comments on the White House Big Data Initiative

Attention: Big Data Study

AMERICAN CIVIL LIBERTIES UNION
WASHINGTON
LEGISLATIVE OFFICE
915 15th STREET, NW, 6TH FL
WASHINGTON, DC 20005
T/202.544.1681
F/202.546.0738
WWW.ACLU.ORG

LAURA W. MURPHY
DIRECTOR

NATIONAL OFFICE
125 BROAD STREET, 18TH FL.
NEW YORK, NY 10004-2400
T/212.549.2500

The American Civil Liberties Union (ACLU) writes today to describe some concrete, immediate steps the Obama administration can pursue to improve privacy and address the expanded use of personal information and big data.[1]

The Big Data Study Group has an enormous task. According to Senior White House Advisor John Podesta, in a mere 90 days, it will "deliver to the President a report that anticipates future technological trends and frames the key questions that the collection, availability, and use of "big data" raise – both for our government, and the nation as a whole."[2]

We commend the Big Data Study Group for the serious and focused attention it has brought to privacy issues in a short period of time. The first two workshops on the issue have been excellent explorations of some of the cutting edge ethical and legal challenges exposed by the accelerating collection of personal information. But we know that any 90 day review will only be a beginning in addressing big data. We also know that big data does not present wholly – or even mostly – new challenges. In reality these issues have been confronting policymakers since at least the 1970s, when the federal government developed the first version of the Fair Information Practice Principles.

[1] The ACLU is a nationwide, non-partisan organization of more than a half-million members, countless additional activists and supporters, and 53 affiliates nationwide dedicated to enforcing the fundamental rights of the Constitution and laws of the United States. The ACLU's Washington Legislative Office (WLO) conducts legislative and administrative advocacy to advance the organization's goal of protecting privacy rights including use of information by government and the private sector.

[2] John Podesta. "Big Data and the Future of Privacy." *The White House Blog.* Jan. 23, 2014 Available at: http://www.whitehouse.gov/blog/2014/01/23/big-data-and-future-privacy

In fact, we already have solutions for some of the privacy issues that confront us today and there are specific actions the executive branch can take to improve Americans' privacy. With that goal in mind, the bulk of these comments will focus on two main areas. The first area is immediate actions the administration can and should take to improve how the federal government collects and uses personal information. The second area is a few specific subjects where sustained focus and attention could improve privacy knowledge and best practices in the future.

A hallmark of the ACLU is the breadth of our work on privacy and our expertise across a wide range of issues, including commercial data collection and use, law enforcement practices, and national security issues. As a final matter, we have prepared as an appendix to this letter a non-exhaustive review of recent ACLU reports and congressional testimony on privacy issues ranging from collection of phone record data by the NSA, to license plate readers, to immigration databases like E-Verify. We hope these will be a valuable resource for exploring specific subjects in more depth.

I. Subjects for Immediate Action

a. Support for Legislation

There is an almost universal acknowledgement that laws related to privacy are out of date. Technology has changed, but the law has not, creating serious gaps in privacy protections or leaving entire areas almost completely unprotected. Obviously Congress, not the Executive, passes new statutes; however there are three areas where the Administration could contribute to the legislative discussion in a way that would advance privacy.

Endorse the USA Freedom Act. Ongoing revelations regarding the use of big data to gather personal information on the American public by the National Security Agency (NSA) and other members of the intelligence community have highlighted the need for reform. The bipartisan USA Freedom Act reins in bulk collection of American records. It amends Section 215 of the Patriot Act – which is used to collect the phone records of almost every American every day – so that it can no longer be used in such a sweeping fashion. The bill would also require individualized suspicion for national security letters and pen registers, two other Patriot Act tools used to access Americans' records. The bill would make changes to the FISA Amendments Act (FAA) to prevent the government from searching through FAA-collected data for U.S. person data in the absence of an emergency or a court order. Finally, the bill includes the creation of a special advocate before the FISA court and new transparency requirements.

Two independent panels – the Privacy and Civil Liberties Oversight Board (PCLOB) and the President's Review Group on Intelligence and Communications Technologies – have already confirmed that intelligence agencies have interpreted their authority in an overbroad and unconstitutional manner. Further these panels "have not identified a single instance involving a threat to the United States in which the [215 telephone records] program made a concrete

difference in the outcome of a counterterrorism investigation."[3] While the President has just called for legislation to better protect Americans' phone records, our email, internet, financial and other records are just as sensitive and require stricter limitations. Endorsement by the Obama Administration of the USA Freedom Act would strengthen reform efforts and confirm that when big data runs amok it must be reined in.

Update the Electronic Communications Privacy Act (ECPA). Among many other protections, ECPA regulates how the government can access the contents of electronic communications. Unfortunately, it has not been substantially updated since 1986. Under the statute, e-mail, documents stored in the cloud, and other private communications like photos and text messages do not receive the protection of a search warrant approved by a judge (the protection that would apply to physical mail or even electronic communications that are not stored with companies like Google or Yahoo).

There is bipartisan legislation in the Senate and the House, S. 607 and H.R. 1852, which would make a simple fix to the law to assure that regardless of where individuals store their communications, those communications will be safe from unjust government intrusion and only accessible with a search warrant based on probable cause. Some areas must be shield from big data analysis without an appropriate legal predicate.

Seemingly, the only major impediment to passage is an objection by the Securities and Exchange Commission, which would like to use the legislation as an opportunity to expand its investigative authority. Support from the Administration would be a major step toward removing this roadblock.

Release Commercial Privacy Model Language. In February 2012 the Administration released a report outlining the need for a "Consumer Privacy Bill of Rights." The report delineated a strong framework for consumer privacy rights, one based on the fair information practice principles and resulting from a multiyear effort to identify and develop workable practices to make those principles a reality. The President committed, "My Administration will work to advance these principles and work with Congress to put them into law."[4]

Unfortunately, more than two years later, there has been no congressional initiative or filed bill. According to press reports, legislative language has been drafted.[5] This language should be

---

[3] Privacy and Civil Liberties Oversight Board. *Report on the Telephone Records Program Conducted under Section 215 of the USA PATRIOT Act and on the Operation of the FISC*. January 23, 2014. Available at: http://www.pclob.gov/SiteAssets/Pages/default/PCLOB-Report-on-the-Telephone-Records-Program.pdf (pg 11)

[4] The White House, *Consumer Data in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Economy.* February 2012. Available at: http://www.whitehouse.gov/sites/default/files/privacy-final.pdf (presidential introduction)

[5] For more information on this problem please see: Alex Byers. "White House pursues online privacy bill amid NSA efforts." *Politico*. Oct. 7, 2013. Available at: http://www.politico.com/story/2013/10/white-house-online-privacy-bill-nsa-efforts-97897.html

released, perhaps as an appendix to the Study Group's report. This language will help to advance the debate, giving activists and privacy supporters concrete reforms they can point to in the fact of continuing privacy invasions.

b. Administrative Reform

Today we live in a world of records. They are generated by electronic devices and stored in massive databases. It is easy to track each of us using cell phones, automated license plate cameras and a host of other technologies. Too often, the result is that the government stores, accesses, and uses personal information on innocent Americans without a reason. Any meaningful regulation of big data must begin by grappling with this reality by bringing transparency to these practices and then regulating or ending their use.

National security surveillance transparency. The long string of revelations regarding US government surveillance both here in the United States and abroad have highlighted how little the American public knows about these programs and the legal authorities that underpin them. It is critical that the administration to be more forthcoming. While we appreciate the President's transparency on the phone metadata program under section 215 of the Patriot Act, there is still a lack of information about the many other surveillance programs currently underway.

Three critical areas in which the administration can advance transparency are:

- release all remaining undisclosed FISA Court opinions;
- describe operational details, scope and legal underpinnings of existing surveillance programs; and
- address how many Americans have had their personal information swept up in these programs and how the government is using the vast amounts of data it is allegedly collecting.

Location and Telephone Record Information. Currently, records related to law enforcement requests for location information and telephone numbers are almost completely secret. In spite of the fact that tens of thousands of these orders are entered annually, Congress, the courts, nor the public has any clear sense of their scope.[6] These orders often reportedly collect not just information on subjects of an investigation, but also on dozens or hundreds of other, completely blameless, individuals.

The Department of Justice should develop a protocol to avoid indefinite sealing of surveillance orders ("D" orders and pen/trap orders). Specifically, a protocol should provide for:

- Immediate review of all sealed applications and orders under 2703(d) and the Pen/Trap statute, followed by DOJ filing motions in district courts seeking unsealing of all applications and orders that do not relate to a currently ongoing investigation or where unsealing will not result in imminent serious risk of physical harm or death to a person;

[6] For more information on this problem please see: Smith, Stephen W., Gagged, Sealed & Delivered: Reforming ECPA's Secret Docket (May 21, 2012). Harvard Law & Policy Review Vol. 6, 2012 Forthcoming. Available at SSRN: http://ssrn.com/abstract=2071399

- Create a prospective DOJ policy requiring US Attorneys' Offices to either seek unsealing of surveillance applications and orders within a reasonable time after an investigation is no longer active, or include a presumptive expiration date for sealing in the applications. For example a seal could expires 180 days after entry of the court's order, unless the government files a motion before that time certifying that the investigation is still active or that unsealing would cause imminent serious physical harm or death.

Automated License Plate Readers (ALPR). Tens of thousands of commercial and government license plate readers collect billions of records on American's location, often keeping that information for years.[7] This information is extremely sensitive, revealing an individual's location at a specific time and potentially where they worship, spend their nights or engage in First Amendment protected activities. Press reports indicate that several federal agencies, including U.S. Immigration and Customs Enforcement (ICE), the Drug Enforcement Administration (DEA) and the Federal Bureau of Investigation (FBI) are building their own ALPR databases or routinely accessing commercial databases.[8] Yet we have few details on how the federal government regulates access to this information.

The two key questions which need to be answered are:

- to what extent are federal agencies building vast databases of ALPR data, and
- to what extent is the federal government accessing data collected by others. These databases include not just private sector data collection but also databases compiled by state/local actors.

Commercial Databrokers. The Privacy Act does not extend to the federal government's use of commercial databases. As documented by a 2008 GAO Report, the federal government uses such databases frequently for a variety of purposes, such as in support of law enforcement and for background check investigations.[9] These databases often contain incorrect information, but individuals currently have none of the protections such as access, notice, correction, and purpose limitations, which are fundamental to the Privacy Act and fair information practices.

Federal agencies should examine and disclose their commercial data access, notice, correction and use policies and perform the same privacy impact assessments (PIAs) on the use of personal information in commercial databases that are already required on agencies' own databases. These PIAs would create basic transparency by requiring agencies to describe what information

[7] For more information see the ACLU's recent report on license plate readers: ACLU. *You Are Being Tracked*. July 2012. Available at: https://www.aclu.org/alpr

[8] Dan Froomkin, "Reports of the Death of a National License-Plate Tracking Database Have Been Greatly Exaggerated." *The Intercept*. March 17, 2014. Available at: https://firstlook.org/theintercept/2014/03/17/1756license-plate-tracking-database/

[9] U.S. Government Accountability Office. (2008, March). Government Use of Data from Information Resellers Could Include Better Protections. (Publication No. GAO-08-543T). Available at: http://www.gao.gov/products/GAO-08-543T

is collected, the purpose of the collection, with whom information will be shared, and how it will be secured.

Surveillance Drones. The federal government increasingly uses unmanned surveillance drones domestically. These small, inexpensive tools have the potential to dramatically increase aerial surveillance and are subject to few legal restrictions. Customs and Border Patrol flies a fleet up to a hundred miles away from both the northern and southern borders. It has also admitted to lending these drones to other federal, state and local law enforcement agencies. According to media reports, such practices have increased eightfold since 2010.[10] Former FBI director Robert Mueller disclosed that the agency uses drones but has yet to develop privacy protocols for that use.[11]

Except in exigent circumstances, agency drone use for criminal investigations should only be conducted pursuant to a particular investigation and after judicial approval. In non-criminal circumstances (such as patrolling the border) drones should not be 'loaned out' or used beyond their stated purpose by the federal agency authorized to use them.

Each of these proposals shares a common idea: that any program that collects data regarding the activities of a substantial number of people for a law enforcement or intelligence purpose, without any individualized suspicion, must be disclosed. When large swaths of people are subject to such collection, fundamental principles of democracy require disclosure so that there can be a public debate about the privacy tradeoffs. That principle should be applied broadly across the federal government.

II. Future Areas of Investigation

As the Study Group focuses on the future impact of big data, it should pay specific attention to two areas – the reality that data collection may exacerbate existing inequality and discrimination and effective research techniques that can be developed to protect privacy while allowing research to flourish.

a. Impact of Big Data on Exacerbating Inequality and Discrimination

The ACLU, along with 13 other civil rights, privacy and media justice organizations, is a signatory to five civil rights principles for the era of big data. These principles recognize the importance of data collection for documenting persistent inequality and discrimination but also seek to build an intellectual framework for assessing how surveillance and data use is being

[10] Craig Whitlock and Craig Timberg, "Border-patrol drones being borrowed by other agencies more often than previously known," *Washington Post*. Jan. 14, 2014. Available at: http://www.washingtonpost.com/world/national-security/border-patrol-drones-being-borrowed-by-other-agencies-more-often-than-previously-known/2014/01/14/5f987af0-7d49-11e3-9556-4a4bf7bcbd84_story.html.

[11] Jake Miller, "FBI Director Acknowledges Domestic Drone Use." *CBS News.* June 19, 2013. Available at: http://www.cbsnews.com/news/fbi-director-acknowledges-domestic-drone-use/

woven into the fabric of ordinary life, sometimes with harmful effects. The specific principles are:

1. Stop High-Tech Profiling.
2. Ensure Fairness in Automated Decisions.
3. Preserve Constitutional Principles.
4. Enhance Individual Control of Personal Information.
5. Protect People from Inaccurate Data.[12]

Data and surveillance are already part of ordinary life, especially in poor or disadvantaged communities that have long faced excessive government scrutiny. For example, a recent news story described an invasive, new police tactic employed by the Chicago Police Department.[13] Using software created by an engineer at the Illinois Institute of Technology, the city developed a "'heat list' — an index of the roughly 400 people in the city of Chicago supposedly most likely to be involved in violent crime." The criteria for placement on the list are secret, but reportedly go beyond indicators like criminal conviction, and raise real questions about racial bias in the selection process.

The results of placement can be very invasive. At least one person reported that a Chicago police commander showed up at his door to let him know the police would be watching him. He hadn't committed a crime or even recently interacted with police. This type of automated profiling is both a privacy problem and a civil rights problem. Use of personal information to make secret determinations is a violation of privacy. When there is significant potential for racial discrimination and police abuse, that's a civil rights problem.

The Chicago list is just the tip of an iceberg of dangerous ways that big data is being used. A Senate Commerce Committee report recently described marketers' use of lists based on racial and other characteristics to identify "the most and least desirable consumers."[14] The government E-Verify database, which many employers check to determine immigration status, has

[12] For a full description of the principles please see: "Civil and Human Rights Orgs Speak Out for the First Time on Privacy and Big Data Policy," Feb. 27, 2014. Available at: http://www.civilrights.org/press/2014/civil-human-rights-orgs-speak-out-on-big-data-privacy.html

[13] Matt Stroud. "The minority report: Chicago's new police computer predicts crimes, but is it racist?" *The Verge.* Feb. 19, 2014. Available at: http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist

[14] U.S. Senate Committee on Commerce, Science, and Transportation, *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes.* December 2013. Available at: http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0d2b3642-6221-4888-a631-08f2f255b577 (pg 25)

a persistent bias that causes legal immigrants to be wrongly identified as ineligible to work.[15] Police too frequently spy on innocent people who pray at mosques.[16]

All of these examples point to a growing need to consider how privacy and civil rights intersect. As one memorable article recently noted, often the best way to predict the future of surveillance is to ask poor communities what they are enduring right now.[17] The study group should consider these examples and the many others which accompany the principles as it assesses its future focus. While technology and computer analytics sometimes appear to be neutral, in fact they too frequently mirror persistent, existing inequality.

b. Research on Gaining From Big Data without Harming Privacy

In addition to the potential for inequality and privacy invasions, the use of big data may also have very real potential value – for improving medicine, combating climate change and addressing a host of societal ills. It is imperative that privacy not be pitted against those values. Data collection and use should not be a zero sum game where the increased value of data means a decrease in privacy. One of the ways to prevent that outcome is by supporting research which allows scientists to make use of data in noninvasive ways.

Computer scientists are developing new ways to share and analyze large datasets while strongly protecting privacy. One basic risk when removing personal information in order to share a sensitive dataset, is that the data might later be combined with outside information to reveal information on individuals. For example, the research of Dr. Latanay Sweeney has demonstrated that de-identified medical records can be combined with publicly available datasets to re-identify particular individuals and their medical conditions. But groundbreaking advances in differential privacy offer new tools to statistically measure and reduce that risk. The Census Bureau has already adopted differential privacy techniques, using them in its OnTheMap project to publish geographical information about where workers live and work without revealing anyone's specific employment.[18]

Other fundamental breakthroughs in cryptographic research are making it possible to reap the benefits of cloud computing while protecting sensitive information. Cloud computing services currently require access to their users' sensitive data, in order to analyze, search and present the

---

[15] ACLU. *The 10 Big Problems with E-Verify*. May 2013. Available at: https://www.aclu.org/10-big-problems-e-verify

[16] Noa Yachot. "With No Evidence of Wrongdoing, NYPD Treats Entire Mosques as Terrorsits Groups." *The ACLU Blog of Rights*. Aug. 28, 2013 Available at: https://www.aclu.org/blog/national-security-religion-belief-technology-and-liberty-criminal-law-reform/no-evidence

[17] Virginia Eubanks. "Want to Predict the Future of Surveillance? Ask Poor Communities" *The American Prospect*. Jan 15, 2014. Available at: http://prospect.org/article/want-predict-future-surveillance-ask-poor-communities

[18] Erica Klarrech. "Privacy by the Numbers: A New Approach to Safeguarding Privacy" *Quanta Magazine.* Dec 10, 2012. Available at: https://www.simonsfoundation.org/quanta/20121210-privacy-by-the-numbers-a-new-approach-to-safeguarding-data/

information. But a new family of approaches called homomorphic encryption may allow cloud providers to offer useful services together with strong privacy — searching and analyzing users' data without decrypting that data. This approach could leave users in control of their own information in a new and technologically robust way, and might also have beneficial legal consequences (because the user's key for decrypting her data may never have to be shared with any third party).

Strong research funding from the federal government will be critical in order to develop these potentially transformative, privacy-strengthening technologies to make them ready for widespread use. Closer collaborations between engineers and the policy community — such as the interactions fostered by this very review — will likewise remain vitally important as this research continues to develop.

We have attempted to offer manageable actions the executive branch can pursue in the short and long term to increase legal protections and transparency, reduce spying on innocent individuals, and address fruitful avenues for future research. The ACLU urges the Study Group to pursue each of these recommendations. For additional questions please contact Chris Calabrese at (202) 715-0839 or ccalabrese@aclu.org.

Sincerely,

Laura W. Murphy
Director, Washington Legislative Office

Christopher Calabrese
Legislative Counsel

Appendix A

- Section 702 of the FISA Amendments Act: Public Hearing Before the Privacy and Civil Liberties Oversight Board: (March 2014) (statement of Jameel Jaffer, Deputy Legal Director of the ACLU) Available at: https://www.aclu.org/sites/default/files/assets/pclob_fisa_sect_702_hearing_-_jameel_jaffer_testimony_-_3-19-14.pdf

- ACLU, U.S. Government Watchlisting: Unfair Process and Devastating Consequences (March 2014). Available at: https://www.aclu.org/sites/default/files/assets/watchlist_briefing_paper_v3.pdf

- Chris Conley, ACLU of Northern California, Metadata: Piecing Together a Privacy Solution (February 2014). Available at: http://www.datascienceassn.org/sites/default/files/Metadata%20Report%202014-%20Piecing%20Together%20a%20Privacy%20Solution.pdf

- The Future of Unmanned Aviation in the U.S. Economy: Safety and Privacy Considerations: Hearing Before the Senate Commerce Committee (January 2014). (statement of Chris Calabrese, Legislative Counsel of the ACLU) Available at: https://www.aclu.org/sites/default/files/assets/domestic_drones_statement_senate_commerce_committee.pdf

- Nicole Ozer and Matt Cagle, ACLU of Northern California, Losing the Spotlight: A Study of California's Shine the Light Law (November 2013) Available at: https://www.aclunc.org/sites/default/files/Losing%20the%20Spotlight%20-%20A%20Study%20of%20California%27s%20Shine%20the%20Light%20Law%20final.pdf

- Jay Stanley, ACLU, Police Body-Mounted Cameras: With Right Policies in Place, A Win For All (October 2013) Available at: https://www.aclu.org/technology-and-liberty/police-body-mounted-cameras-right-policies-place-win-all

- Jennie Pasquarella, ACLU of Southern California, Muslims Need Not Apply (August 2013). Available at: http://www.aclusocal.org/CARRP/

- Catherine Crump, ACLU, You Are Being Tracked: How License Plate Readers Are Being Used to Record Americans' Movements (July 2013). Available at: https://www.aclu.org/technology-and-liberty/you-are-being-tracked-how-license-plate-readers-are-being-used-record

- Strengthening Privacy Rights and National Security: Oversight of FISA Surveillance Programs: Hearing Before the Senate Judiciary Committee (July 2013) (statement of Jameel Jaffer, Deputy Legal Director of the ACLU and Laura Murphy, Director of the ACLU's Washington Legislative Office) Available at: https://www.aclu.org/files/assets/testimony.sjc_.073113.final_.pdf

- Jay Stanley, Senior Policy Analyst, ACLU and Chris Calabrese, Legislative Counsel, ACLU, Prove Yourself to Work: The 10 Big Problems with E-Verify (May 2013) Available at: https://www.aclu.org/files/assets/everify_white_paper.pdf

- State of Federal Privacy and Data Security Law: Lagging Behind the Times?: Hearing Before the Senate Committee on Homeland Security Subcommittee on Oversight (July 2012) (statement of Chris Calabrese, Legislative Counsel of the ACLU) Available at: https://www.privacysos.org/sites/all/files/Calabrese.pdf