Before the
White House Office of Science and Technology Policy
Washington, D.C.

In the matter of	)
Government Big Data	)
	)
	)

COMMENTS OF
THE COMPUTER AND COMMUNICATIONS INDUSTRY
ASSOCIATION

1 Introduction

In response to the Request for Information released by the Office of Science and Technology Policy (OSTP) on March 4, 2014,[1] the Computer & Communications Industry Association (CCIA) offers the following comments.

CCIA is an international, nonprofit association representing a broad cross section of computer, communications and Internet industry firms. CCIA remains dedicated, as it has for over 40 years, to promoting innovation and preserving full, fair and open competition throughout our industry. Our members employ more than 600,000 workers and generate annual revenues in excess of $200 billion.[2]

These comments will address four issues: 1) That big data is simply a tool and is not uniform in its implications. Its implications depend on who it is using it and for what; 2) Commercial use of data creates value and avoids harm when users are protected through transparency and control; 3) Many government uses of large data sets are vital for the operation of our country, and the Privacy Act protects the subjects of the data, and; 4) That big data analysis for surveillance purposes is highly risky because of the relative

[1] Government "Big Data"; Request for Information, 79 Fed. Reg. 12251 (Mar. 4, 2014).

[2] A complete list of CCIA members is available at http://www.ccianet.org/members/.

power of the state in law enforcement and national security contexts, and that therefore stringent restrictions and oversight is necessary.

2 Big data is simply a tool

The term "big data" has become a shorthand for an enormous variety of technologies and techniques for the gathering, manipulation, and analysis of large data sets. It is important in a policy examination of big data to realize that the term does not encompass a monolith of practices. Indeed, the subject pertains to a diverse set of tools, and in reality there is still considerable disagreement over what precisely the term refers to.[3]

As OSTP proceeds in examining the implications of big data, it will be important to keep this fact in mind. Questions or solutions that appear to apply to the entire waterfront of big data should be examined closely to make sure that they are not actually applicable to only a subset of "big data", or applicable to different areas in different ways.

In particular, government uses of big data are different from commercial ones for a variety of reasons. Solutions that protect data subjects while promoting innovation and product solving in one situation may be inapplicable or even harmful in another. To the extent that OSTP makes recommendations, they will be more helpful if this fact is kept in mind.

Even within the category of commercial uses of big data, there will be variation in what kinds of data is collected, whether the data is individually marked or aggregated, and the uses that the data is put to. In many cases, some of these factors will inherently protect users from real harm while in others further interaction will be necessary. In any case, OSTP should be conscious of these distinctions if attempting to make broad observations.

When approaching such a broad and complex subject, it is wise to approach the problem with an awareness of and sensitivity to the unknown. OSTP has already demonstrated considerable sensitivity to the as-yet undis-

[3] *See, e.g., The Big Data Conundrum: How To Define It*, MIT Technology Review, Oct. 3, 2013, http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/.

covered technological innovations in the big data area by calling upon the public to advise them through this proceeding and other meetings. Applying that same awareness to any resulting recommendations will ensure that they are relevant and proportional to the issues they are addressing.

3 Commercial applications of big data bring value to users and offer proper controls

The commercial uses of "big data" are as numerous as there are business plans in the country. The basic analysis of the data produced by a company can be considered big data and can contribute to significant consumer benefits, increasing sales, greater efficiencies, streamlining supply chains, and an innumerable host of other applications. In the years to come more and more companies will be exploring the data they generate as a routine part of doing business. Many of these uses will have no privacy implications whatsoever.

There are a number of reasons why privacy may not be implicated. In situations where a company is simply analyzing data about their own operations, such as supply chain information, there is simply no personal information about consumers involved and thus no privacy impact. In other circumstances, a company may be working with data about individuals who are customers of the company, in order to improve services, and where the data never leaves the company itself. In this case the privacy implications are minimal. Aggregated or deidentified data may also pose considerably less of a privacy concern, as data on particular people is not available. This is even more true if the company commits to not attempting to reidentify the data in the course of their work with it.

There are also many cases where privacy interests are implicated. If data collected and analyzed is about identified individuals, and there is a potential for concrete harm to those individuals, then privacy concerns are at their apex. In those cases there are a few ways in which people can be protected from harm. First, there are some areas where the concrete harm can be so great that we have enacted laws to control the use of the data, and

our current legal framework is fully capable of providing remedies for these harms. For example, in situations where particular adverse decisions may be made against a person from some data, the Fair Credit Reporting Act controls how the data may be used and what rights the person has to see and correct information about them.[4] Secondly, where no law applies to the data collection and analysis, it is important that companies be transparent about their actions and give users control over how the data is used, including the opportunity to correct inaccurate data and the possibility of opting-out if they desire. In these circumstances there is always the backstop of the Federal Trade Commission, using its authority to make sure that companies follow the public pledges they make about privacy.

Commercial big data products are important because they make possible a huge variety of features and products that benefit consumers, and even entire business plans. Mobile mapping applications that warn about traffic do so by aggregating huge amounts of location data from phones traveling on the roads. Apple's voice recognition software only works from analyzing a large corpus of voice recordings and refines itself all the time based on the queries it receives from users. Self-driving cars, like those being developed by Google and other companies, rely on having detailed information on the roads that the car will be driving along. While not a perfect crystal ball, large-scale data analysis can also create insights that lead to the next great American company, not just in Silicon Valley, but around the country.

Two years ago the White House proposed a Consumer Privacy Bill of Rights, a framework that was intended to capture common privacy principles in a "comprehensive" way. When it released the Bill of Rights, the White House appropriately praised the strength of the U.S. privacy regime:

> The consumer data privacy framework in the United States is, in fact, strong. This framework rests on fundamental privacy values, flexible and adaptable common law protections and consumer protection statutes, Federal Trade Commission (FTC) enforcement, and policy development that involves a broad array

[4]Fair Credit Reporting Act, 15 U.S.C. §1681 et seq.

> of stakeholders. This framework has encouraged not only social and economic innovations based on the Internet but also vibrant discussions of how to protect privacy in a networked society involving civil society, industry, academia, and the government.[5]

The 2012 report aimed to address concerns raised by those who favored a single approach applying across all sectors by articulating "a clear statement of basic privacy principles" and catalyzing "a sustained commitment of all stakeholders to address consumer data privacy issues as they arise from advances in technologies and business models."

The privacy world has changed considerably since the Administration's 2012 Report, during which time the U.S. has "experimented with new methods of privacy policymaking in the form of NTIAs multistakeholder process. And in corporate America, a culture of privacy awareness has blossomed into robust privacy programs and a thriving market for privacy professionals."[6] In contrast, enforcement of "comprehensive" privacy regulation has been more stagnant and limited, and debate continues about how to update European privacy laws for the twenty-first century.

One final feature of big data for OSTP to keep in mind with regard to commercial uses is the way it can quickly change the landscape of a marketplace, including its own. The inexpensive availability of data analysis, extending to nearly all actors in a marketplace, can cause rapid shifts in business methods and products. In the face of such a quickly moving set of circumstances, regulators must be careful when trying to address issues. Any government-imposed solution risks being outdated before it is even implemented. Multistakeholder processes that involve consumer groups, industry, and government working together are better equipped to tackle privacy issues arising from big data in flexible but privacy protecting ways. Legislation does not easily so adapt to the ever-evolving nature of commercial

[5] The White House, Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy i (2012).

[6] *See* Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy in Europe: Initial Data on Governance Choices and Corporate Practices*, 81 Geo. Wash. L. Rev. 1529, 1564-65 (2013).

technologies in the same way.

4 Government is beginning to explore big data

There are a number of areas in which government is also beginning to see how big data can make the business of running government more effective and efficient, as well as opening up services that previously would have been impossible. While the techniques of analysis that come with the big data moniker are new to government, the government has always been a major collector of data, at large scale, and of highly personal nature. Of course government must therefore focus on the privacy of people whose data is being analyzed, and the Privacy Act provides the rules for how that happens.

The ways in which the government is using big data are as varied as the government agencies themselves. The Library of Congress is cataloging every public tweet ever published for future analysis or even just posterity. The Center for Disease Control employs statistics and big data to track flu outbreaks.[7] The various statistical agencies, such as the Bureau of Justice Statistics and the Bureau of Labor Statistics, have begun releasing data in machine readable formats, allowing them to be combined with other data sets for greater depth of analysis.[8] All of these efforts show how the power of big data is influencing how government operates.

The privacy of data subjects when the government collects and analyzes large amounts of data is no less important just because the government is the one doing the collecting. Indeed, in some ways, it is more important, as residents of a country have little opportunity to simply select a different government to live under, while users of online services can usually easily pick and choose a service that has privacy options they prefer. Competition for users in the online world provides a form of "market discipline" on companies and how they use data. Because governments are not subject to the same changes in market dynamics, they are not as responsive to privacy

[7] Note that CDC's flu tracker is separate from Google's and studies have shown that the most accurate representation of real-world flu comes from combining the two approaches. *See* Steve Lohr, *Google Flu Trends: The Limits of Big Data*, N.Y. Times, Mar. 28, 2014.

[8] *See generally*, Bureau of Justice Statistics, http://www.bjs.gov.

problems and their feedback mechanisms break easily. For these reasons, protections on government use and analysis are important.

The law that controls how the government treats the data it collects and analyses is the Privacy Act of 1974.[9] The Privacy Act was one of the first attempts to deal with the idea of large computer databases, operated by the government, collecting information about citizens. The act grew out of a process at the Department of Health, Education, and Welfare that codified what are now known as Fair Information Practice Principles (FIPPs). Those principles were adapted into the Privacy Act and today control how the government treats data.

The Privacy Act, however, has not been substantially modified since its passage in 1974, nor has the guidance issued by the Office of Management and Budget in 1975. The intervening 40 years, on the other hand, have seen an explosion of new technology and new applications of that technology to data analysis. For example, the Privacy Act's central definition, a "system of records," is hard to accurately apply in a modern age where many interlocking databases may be maintained, sometimes by different agencies.[10] Similarly, the "routine use" exception to the Privacy Act is poorly defined and in practice operates as a huge loophole for government agencies to share personal information, and could be a focal point for reform.

There have been occasional efforts to update the law, but none have succeeded. OSTP's exploration of the implications of big data should focus on whether the Privacy Act is adequate to address the federal government's data usage today. If the White House recommends upgrades to the Privacy Act, it could catalyze what has been a stalled conversation about the direction of the law in the 21st Century.

[9] Privacy Act, 5 U.S.C. §522a et. seq.

[10] *See, e.g., Veterans Data Breach Highlights Inadequate Privacy Protections*, Center for Democracy and Technology, May 31, 2006, available at https://www.cdt.org/policy/veterans-data-breach-highlights-inadequate-privacy-protections.

5 The use of big data for surveillance is inherently problematic

While normal government use of big data can be troublesome in some contexts (making the Privacy Act an important bulwark against abuse), big data for surveillance purposes, either in a criminal context or for national security, brings in a host of problems. One essential solution available today is for Congress to pass and the President to sign a bill that updates the Electronic Communications Privacy Act of 1986 (ECPA) to ensure a warrant for content standard.[11] In the national security context, there are a series of much-needed reforms to how the government treats the vast amounts of information it collects in the name of security. In particular, the recent decisions with regard to bulk collection of metadata are a good start, but there are still problems to be addressed.

Data collected under surveillance regimes are different from all the other forms of data. Out of all of these categories, the potential harms that can come from abuse are greatest from government surveillance. That is why we have historically had such strong controls on the government's ability to gather information for the purposes of criminal investigation. Those controls were enshrined in the Fourth Amendment to our Constitution and have been, for most of the history of the Republic, the primary protection against abuse by the government.

Today that protection is no longer adequate. More and more of our day to day lives are now spent online. Communication, work, play, and creative pursuits all have moved to the Internet in some form or another. The law, sadly, has not kept up. Beginning with the Third Party Doctrine created by the Supreme Court in Smith v. Maryland in 1979, a distinction has been made between the lives we live offline and the lives we live online.[12] Paper mail is protected while email is not. Files in a file cabinet are protected while files in cloud storage are not. For most Americans today, however, these distinctions are meaningless.

[11] Electronic Communications Privacy Act, 18 U.S.C. §2510 et. seq.
[12] Smith v. Maryland, 422 U.S. 735 (1979).

Fortunately, the solution already exists. A clean bipartisan and bicameral bill exists that would fix the warrant for content problem in a narrow and targeted way. The Senate version, S. 607 (written by Senator Leahy, the original author of ECPA), has already passed the Senate Judiciary Committee on a voice vote, and the House version, H.R. 1852, now has almost 200 co-sponsors from both sides of the aisle.[13] The White House could help make this easy fix a reality, improving Americans' privacy, giving certainty to companies doing business online, and showing the international community that responsible control over government surveillance is a priority of the U.S. government.

In the national security arena, there is still a lot that the White House should do to bring the National Security Agency (NSA)'s practices into line with principles that will protect people around the world and encourage trust in the online marketplace. CCIA supports the Reform Government Surveillance principles promulgated by many of the large tech firms.[14] These principles deal with the concept of big data directly and should serve as guideposts for the administration as it considers how it will move forward with its national security work.

The first principle, "Limiting Governments' Authority to Collect Users' Information," calls for governments to target surveillance at specific known users and only for lawful purposes. By limiting collection to this category, this principle seeks to take national security surveillance out of the category of big data entirely. Indiscriminate gathering of information about the public, in order to sift it for possible indications of wrongdoing, is not compatible with this idea.

With regard to particularity of government data collection, the President's recent announcement of restrictions on the bulk collection of telephone metadata is very welcome, however there are two remaining issues that the government should address. First, the announcement applies to telephone metadata only at the moment. While it appears as if the NSA

[13] Electronic Communications Privacy Act Amendments Act, S. 607, 113th Cong. (2013); Email Privacy Act, H.R. 1852, 113th Cong. (2013).

[14] *See* Reform Government Surveillance, https://www.reformgovernmentsurveillance.com/.

is not currently gathering metadata from online transactions,[15] under the plan proposed by the President it could resume at any time. The new rule should apply to the bulk collection of any metadata, no matter where it lives. Secondly, the new rule still maintains the troubling problem of permitting the NSA to chain "hops" of people who communicate together to reach non-targets. This is problematic particularly when one of the hops is a phone number that a large number of people call, such as a pizza delivery number or a government service such as a Department of Motor Vehicles.

The second and third principles cover the equally important question of what happens surrounding the collection of data. Governments should place their collection apparatuses under proper oversight and must hold accountable the groups doing the collection.

The fifth principle, "Avoiding Conflicts Among Governments," also pertains to big data in surveillance. Mutual Legal Assistance Treaties (MLATs) are structures by which the investigating authorities in one country can obtain information from companies located in another country. The government of the requesting country makes the demand of the government where the company is located and the second government then makes the demand of the company. This process, when it works, gets investigators the information they need, while at the same time forcing them to follow proper channels (rather than simply attempting to intimidate any local staff of the target company into turning over information on risk of imprisonment).

The MLAT process, however, almost never works as it should. The process is convoluted and inefficient. Requests can take up to 18 months to process in some cases.[16] Fortunately, the administration has signalled its intention to fix it. The President in January announced his intention to reform the MLAT process and the Department of Justice has asked Congress

[15]Glenn Greenwald & Spencer Ackerman, *NSA collected US email records in bulk for two years under Obama*, The Guardian, June 27, 2013, available at http://www.theguardian.com/world/2013/jun/27/nsa-data-mining-authorised-obama.

[16]The President's Review Group on Intelligence and Communications Technologies reported an average of 10 months backlog and some cases that go back substantially further than that. *See*, *e.g.*, The President's Review Group on Intelligence and Communications Technologies, Liberty and Security in a Changing World (2013) at 227.

for a larger budget to help hire more lawyers and staff to deal with the MLAT backlog.[17]

While those are good first steps, there are still other things that the U.S. government could be doing to improve the process. As in most other areas of surveillance, transparency about the MLAT process should be a priority. Only through knowing how the current process is working can we effectively attempt to fix it. Transparency will also show where the most requests come from and which countries are using the process, and help therefore highlight where the most effective reforms might take place.

6 Conclusion

The White House, in exploring the issues that are brought about by big data, should keep in mind that uses of big data may take many forms and have many different implications. Attempting to address all of them in the same way will be ineffective. In particular, OSTP should be conscious of the different categories outlined in these comments and think about how those categories pose different issues. Finally, because of the immense harm that can arise, OSTP should be most critical of the use of big data for surveillance purposes.

CCIA thanks OSTP for the opportunity to comment on this important matter, and would welcome the opportunity to answer any questions or make any clarifications that are requested.

[17] See Press Release, U.S. Department of Justice, Attorney General Holder Announces President Obama's Budget Proposes $173 Million for Criminal Justice Reform (Mar. 4, 2014), available at http://www.justice.gov/opa/pr/2014/March/14-ag-224.html.