BIG DATA:

SEIZING OPPORTUNITIES, PRESERVING VALUES

Interim Progress Report

February 2015

One year ago, President Obama spoke at the Department of Justice about changes in the technology we use for national security and signals intelligence purposes, and what those technological changes mean for privacy writ large. Recognizing that these technologies have implications beyond the national security arena, the President also called for a wide-ranging review of big data and privacy to explore how these technologies are changing our economy, our government, and our society, and to consider their implications for personal privacy. The goal of the review was to understand what is genuinely new and different about big data and to consider how best to encourage the potential of these technologies while minimizing risks to privacy, fair treatment, and other core American values.

Over the course of the 90-day inquiry, the big data and privacy working group—led by Counselor to the President John Podesta, Commerce Secretary Penny Pritzker, Energy Secretary Ernest Moniz, the President's science advisor Dr. John Holdren, and the President's economic advisor Jeff Zients—sought public input and engaged with academic researchers and privacy advocates, regulators and the technology industry, and advertisers and civil rights groups. The review was supported by a parallel effort by the President's Council of Advisors on Science and Technology (PCAST) to investigate the scientific and technological dimensions of big data and privacy.

The big data and privacy working group's report found that the declining cost of data collection, storage, and processing, coupled with new sources of data from sensors, cameras, and geospatial technologies, means that we live in a world where data collection is nearly

ubiquitous, where data retention can be functionally permanent, and where data analysis is increasingly conducted in speeds approaching real time. While there are promising technological means to better protect privacy in a big data world, the report's authors concluded these methods are far from perfect, and technology alone cannot protect privacy absent strong social norms and a responsive policy and legal framework. Finally, the report raised issues around other values potentially implicated by big data technology—particularly with regard to the potential for big data technologies to lead, purposely or inadvertently, to discriminatory outcomes on the basis of race, gender, socioeconomic status, or other categories.

But big data technologies continue to hold enormous promise, as the report identified—to streamline public services, to advance health care and education, and to combat fraud and complex crimes like human trafficking. A year after the President's request for this report, the Obama Administration has worked to advance a number of the concrete policy proposals offered in the report, both by launching new efforts and continuing to develop previously existing projects. The Administration continues to drive the national conversation, inside and outside of government, on how to maximize benefits while minimizing the risks and harms posed by a big data world.

Key Recommendations

The big data and privacy working group report identified six specific policy recommendations as deserving prompt action:

- **Advance the Consumer Privacy Bill of Rights** because consumers deserve clear, understandable, reasonable standards for how their personal information is used in the big data era.
- **Pass National Data Breach Legislation** that provides for a single national data breach standard, along the lines of the Administration's 2011 Cybersecurity legislative proposal.
- **Extend Privacy Protections to non-U.S. Persons** because privacy is a worldwide value, and should be reflected in how the federal government handles personally identifiable information from non-U.S. citizens.
- **Ensure Data Collected on Students in School is used for Educational Purposes** to protect students from having their data shared or used inappropriately.
- **Expand Technical Expertise to Stop Discrimination** so that the federal government's lead civil rights and consumer protection agencies can identify practices and outcomes facilitated by big data analytics that have a discriminatory impact on protected classes, and develop plans for investigating and resolving violations of law.
- **Amend the Electronic Communications Privacy Act** to ensure the standard of protection for online, digital content is consistent with that afforded in the physical world—including by removing archaic distinctions between email left unread or over a certain age.

The Administration is making significant progress on most of these recommendations:

- The Department of Commerce solicited public comment on the Consumer Privacy Bill of Rights in light of new technologies, including those identified in the big data and privacy report, and the Obama Administration will release draft legislation in early 2015.

- President Obama released revised national data breach legislation, the Personal Data Notification & Protection Act, on January 12, 2015.

- Attorney General Eric Holder announced in June 2014 that the Administration would seek legislation extending to EU citizens the same right to judicial redress for intentional or willful wrongful disclosure of personal data exchanged under the U.S.-EU Data Protection and Privacy Agreement for law enforcement purposes, or for refusal to grant access or to rectify any errors in that information, as U.S. citizens would have under the Privacy Act of 1974. The Office of Management and Budget is working with departments and agencies to extend other privacy protections to non-U.S. citizens.

- President Obama announced the Student Digital Privacy Act, a national effort to ensure K-12 student data is used only for educational purposes, on January 12, 2015, in conjunction with new private sector commitments to help enhance privacy for students as well as a landmark voluntary effort by over 100 companies committing not to abuse education data.

- Several efforts have been undertaken to further the federal government's understanding of big data and discrimination, including studying the potential implications of using predictive analytics in law enforcement at the Department of Justice and by studying price discrimination at the Council of Economic Advisers. The White House Domestic Policy Council is preparing a follow-on report for release in early 2015 focusing on the potential of big data both to lead to discriminatory outcomes in key policy areas and to be used to counteract discrimination.

Further progress on implementing the big data and privacy report's recommendations and related efforts is detailed in the following pages.

1. Preserving Privacy Values

The innovation driven by big data creates both tremendous opportunity and novel privacy challenges. The report explored privacy challenges across sectors, and suggested that we reexamine our conception of notice and consent, as well as the notion of use frameworks as a basis for managing privacy rights. The report suggested a number of specific steps forward in order to ensure that privacy protections evolve in a way that enables the social good that can result from big data, while protecting and empowering citizens.

Advance the Consumer Privacy Bill of Rights

The report called on the Department of Commerce to advance the 2012 Consumer Privacy Bill of Rights by seeking public comment on big data developments and how they impact the CPBR's policies and then devise draft legislative text. This month, the Administration plans to release draft legislation based on public comments received during that comment period.

Pass National Data Breach Legislation

The report called for the creation of a national data breach standard to benefit both consumers and businesses, in the face of a growing number of breaches and an inconsistent patchwork of state laws. In January 2015, President Obama announced the Personal Data Notification & Protection Act, a new legislative proposal to help bring peace of mind to all Americans, including the tens of millions whose personal and financial information has been compromised in a data breach. This proposal clarifies and strengthens the obligations companies have to notify customers when their personal information has been compromised, while providing companies with the certainty of a single, national standard—as well as criminalizing the illicit overseas trade in identities.

Bring Greater Transparency to the Data Services Industry

In May, the Federal Trade Commission released an in-depth report on the data broker industry, concluding that data brokers operate with a fundamental lack of transparency. The Commission recommended that Congress consider enacting legislation to make data broker practices more visible to consumers and to give consumers greater control over the personal information collected and shared by data brokers.

Lead International Conversations on Big Data

Data privacy has long been a component of the United States' bilateral and multilateral discussions. Well before the big data and privacy report, the Administration engaged in extensive consultation with data protection authorities, international civil society, and privacy experts from Europe and around the world.

Of particular note since the release of the report, high-ranking officials from the United States and Germany discussed the report's findings and bilateral cooperation on cyber issues as part of the third Cyber Bilateral Meeting in June 2014, including cybersecurity and critical infrastructure protection, cyber defense, combating cybercrime, Internet freedom, and Internet governance.

Extend Privacy Protections to non-U.S. Persons

The report recommended that the OMB work with agencies to apply the Privacy Act to non-U.S. persons where practicable, or establish alternative privacy policies for personal data held by the federal government that provide appropriate and meaningful protections regardless of nationality. OMB has been leading an interagency process to implement this recommendation.

In addition to these general protections, the United States is actively pursuing efforts to grant certain rights of judicial redress to EU citizens and citizens of other nations that effectively share terrorism and law enforcement information with the United States and provide appropriate privacy protections. In the 2014 U.S.-EU Ministerial Meeting on Justice and Home Affairs, Attorney General Eric Holder made clear the United States' commitment to pursue this effort, and the Administration is working closely with members of Congress on this important measure.

2. Responsible Educational Innovation in the Digital Age

"[D]ata collected on students in the classroom should only be used for educational purposes – to teach our children, not to market to our children. We want to prevent companies from selling student data to third parties for purposes other than education. We want to prevent any kind of profiling that puts certain students at a disadvantage as they go through school."

- President Barack Obama at the Federal Trade Commission, January 12, 2015

Big data has the potential to transform education for the better, creating unprecedented educational opportunities—for instance, by tailoring lessons to a student's learning style, by opening up courses through online platforms, and by making it easier for parents, teachers, and students to identify where an individual student may be struggling and offer targeted instruction. These new technologies hold the potential to vastly improve student performance and to provide researchers with valuable insights about how students learn, which could help improve low-tech educational interventions as well. Beyond educational technology, the mere operation of schools produces vast amounts of data—data that can improve efficiency as well as education. However, the federal government must play its part to ensure that student data is not shared or used inappropriately. The Administration has taken significant steps to safeguard student data in the classroom and beyond, as well as promoting and enabling innovation in learning.

Ensure Data Collected on Students in School is used for Educational Purposes

On January 12, 2015, the President proposed the Student Digital Privacy Act: a new legislative proposal designed to provide teachers and parents the confidence they need to enhance teaching and learning with the best technology—by ensuring that data collected in the educational context is used only for educational purposes. This bill, modeled on a landmark California statute, builds on the recommendations of the report, would prevent companies from selling student data to third parties for purposes unrelated to the educational mission and from engaging in targeted advertising to students based on data collected in school—while still permitting important research initiatives to improve student learning outcomes, and efforts by companies to continuously improve the effectiveness of their learning technology products.

The legislation will be accompanied by new tools from the Department of Education to empower educators around the country. The Department of Education and its Privacy Technical Assurance Center play a critical role in protecting American children from invasions of privacy in the classroom. Alongside the President's call for legislation, he unveiled executive actions that will enhance that office's abilities to help ensure educational data is used in ways appropriate and in accordance with the educational mission—including a model terms of service and providing teacher training assistance.

The largest educational technology vendors also committed to help lead the way in ensuring the protection of students—and, as of today, over 100 of them have signed on to a pledge to provide important protections against misuse of students' data.

Recognize Digital Literacy as an Important 21st Century Skill

Knowledge and efficient use of digital materials will become increasingly important as computer technologies begin to drive economic and educational empowerment. This recommendation was included in both the big data and privacy working group's recommendations and in the PCAST report. The Administration has advanced several initiatives that encourage digital literacy by connecting Americans to the latest technologies and strengthening the technical skills that can enable fluid use of the latest digital resources. These initiatives promote: (1) the literacy to help students be creators–not just consumers–with increased access to coding experiences, as the President illustrated by participating in the Hour of Code in fall 2014; (2) the literacy to be prepared to work in the STEM fields, through initiatives such as the President's Educate to Innovate campaign; (3) the literacy to use technology smartly, including empowering students to protect their privacy; and (4) literacy realized as access for all, including access to broadband at home and at school, an issue the President has tackled through the ConnectED Initiative. Connectivity is especially critical, as these initiatives must help bridge the digital divide and inequality of opportunity that often exists in educational contexts throughout the nation.

In the coming months, the White House will continue to work with stakeholders and other partners to develop new initiatives to make digital literacy opportunities more accessible and available for the American people.

3. Big Data and Discrimination

One of the most notable findings of the big data and privacy report was that alongside its potential benefits to be used to increase access to credit or improve educational outcomes, there also exists the potential for big data technology to be used to discriminate against individuals, whether intentionally or inadvertently, potentially enabling discriminating outcomes, reducing opportunities and choices available to them.

As part of the national discussion prompted by the big data study, the civil rights community, industry and federal agencies began to identify possible principles and frameworks to guide uses of data. Before the report was completed, a coalition of civil rights organizations announced a set of civil rights principles for the big data era, focused on stopping high-tech profiling, ensuring fairness in automated decisions, preserving constitutional principles, enhancing individual control

of personal information, and protecting people from inaccurate data. The civil rights community worked with technologists and academics to organize an October 2014 conference on big data and discrimination and hopes to make it an annual event, with continuing strong participation from the federal government.

The White House considers this topic a priority, and is continuing to explore the implications of big data in this arena, including considering how big data technology can be used to shore up civil rights. Among other investments, the Obama Administration's budget for Fiscal Year 16 includes $17 million for data science pilots at the National Science Foundation that seek to study issues around data interoperability; data policy and governance; and data security, privacy, integrity, and trustworthiness. These pilots will directly inform other federal big data research projects and will assist in developing the technological and policy expertise needed to tackle difficult problems like the potential for big data to lead to discriminatory outcomes.

Pay Attention to the Potential for Big Data to Facilitate Discrimination

The White House Domestic Policy Council and the Office of Science and Technology Policy will issue a follow-up report further exploring the implications of big data technologies for discrimination and civil rights. Specifically, the new report will take a deeper dive into how big data interacts with issues like employment and access to credit—considering both how the use of big data technologies can perpetuate discrimination and prevent it. The White House has engaged with leading researchers and advocates to develop recommendations on actions that can be taken to use big data to broaden opportunity and to prevent discrimination.

Expand Technical Expertise to Stop Discrimination

One of the key recommendations of the big data and privacy report was that the federal government's lead civil rights and consumer protection agencies should expand their technical expertise to be able to identify practices and outcomes facilitated by big data analytics that may have a discriminatory impact on protected classes, and develop a plan for investigating and resolving potential violations of law.

In June, the Office of Science and Technology Policy and the Georgetown University McCourt School of Public Policy's Massive Data Institute cohosted a fourth big data convening focused on the work of federal agencies. The multi-stakeholder workshop focused on federal agencies' use of open data and big data, best practices for sharing data within and between agencies and other partners, and how to address potential privacy and civil liberties concerns that arise from the use of big data.

In September, the Federal Trade Commission hosted a workshop entitled "Big Data: A Tool for Inclusion or Exclusion?" in its Washington offices. The workshop explored the use of big data and its impact on American consumers, with an eye towards low income and underserved consumers. The workshop highlighted concerns about whether big data may be used to categorize consumers in ways that may affect them unfairly, or even unlawfully.

Deepen Understanding of Differential Pricing

The White House Council of Economic Advisors conducted a study of commercial applications of

big data. The CEA explored whether companies will use the information they harvest to more effectively charge different prices to different customers. The economic literature on value-based price discrimination suggests that this will often, though not always, be welfare-enhancing for both businesses and consumers. However, individualized pricing based on estimates of cost or riskiness can raise concerns about fairness, particularly when consumers are unaware of the data or methods that companies employ. The CEA report finds that many companies already use big data for targeted marketing, and some are experimenting with personalized pricing, though examples of personalized pricing remain fairly limited.

4. Law Enforcement and Security

Big data can be used to make our communities safer and strengthen our national security, but raises equally important questions for our personal privacy and civil liberties. The big data and privacy report encouraged our national security, homeland security, law enforcement, and intelligence communities to vigorously experiment with and employ lawful big data technology while adhering to full accountability, oversight, and relevant privacy requirements.

Review Law Enforcement's Use of Predictive Analytics

In light of the report, the Department of Justice recently conducted a review of the current use of predictive analytics in law enforcement. This review focused on the DOJ's own use of analytic tools, as well as on some of the programs the Department helps fund through research grants. DOJ also reviewed some of the newer technologies in use by state and local law enforcement agencies.

DOJ concluded that new data-driven technologies have the potential to bring significant benefits to our criminal justice system. Many of these technologies build on traditional techniques and are designed to help law enforcement agencies allocate scarce resources more efficiently to prevent crime. The Department also observed that the use of predictive analytics raises issues and potential challenges that are worthy of continued attention, so that predictive techniques continue to be driven by the core enforcement goals of protecting the public and ensuring fairness in our justice system.

Going forward, DOJ will work collaboratively with stakeholders and develop guidance for the use of predictive analytics by state and local law enforcement agencies. The Department will also continue to engage in ongoing conversations about the effectiveness and impact of new predictive techniques.

Foster Responsible Use and Privacy Best Practices with State and Local Law Enforcement Entities Receiving Federal Grants

The big data and privacy report recommended that that federal agencies with expertise in privacy and data practices provide technical assistance to state, local, and other federal law enforcement agencies seeking to deploy big data techniques. In November 2014, DOJ developed a supplemental guide to augment its privacy-related technical assistance library for state, local, and tribal law enforcement agencies, entitled *Resource Guide for Enhancing Community*

Relationships and Protecting Privacy and Constitutional Rights. This supplemental guide serves as a point of reference for state, local, and tribal law enforcement entities in fostering the development of responsible privacy practices. Additionally, DOJ continues to engage in outreach to state, local, and tribal law enforcement entities through participation in trainings and conferences on related issues.

Review Government Use of Commercial Databases

The report recommended that the federal government review uses of commercially available databases on U.S. citizens, focusing on use of services that employ big data techniques and ensuring that they incorporate appropriate oversight and protections for privacy and civil liberties. DOJ and the Office of the Director of National Intelligence, together with the Office of Management and Budget, are leading an effort to review the use of commercial databases by the federal government. In particular, they are examining the use of commercial databases by federal agencies in the context of public administration, law enforcement, and national security. The review process will include recommendations for how the government can use the databases while also protecting privacy and civil liberties.

Implement Best Practices for Controlled Use and Storage of Data at Agencies

Efforts are underway on several fronts to maximize privacy protections by improving agency use and storage of data, and to strengthen cybersecurity in general. For instance, the Department of Homeland Security is working across government and the private sector to identify and leverage the opportunities big data analytics presents to strengthen cybersecurity. This will include coordinating the development or changes of necessary policies to ensure that data is appropriately protected and secured.

The Office of Management and Budget is leading an effort to expand successful data management and security pilots across government and has connected practitioners and leaders from innovative and effective data management initiatives at several federal agencies to foster an exchange of success stories and lessons learned.

The National Security Council is asking the President's National Security Telecommunications Advisory Committee to undertake a private sector-led study with recommendations on using big data analytics to strengthen cybersecurity.

The Administration has also continued to address the challenges to information sharing. The Department of Justice and the Federal Trade Commission issued guidance that sharing of cyber threat information should not raise anti-trust concerns—thus addressing a long-standing concern from industry. The Department of Homeland Security is modernizing its Protected Critical Infrastructure Information program to enable its use for the protection of private sector information voluntarily submitted to the Department for the purposes of improving network defenses.

Advance Cybersecurity and Consumer Protection with 2015 Summit

On February 13, 2015, the White House will host a cybersecurity and consumer protection summit at Stanford University. The summit will bring together major stakeholders on cybersecurity and consumer financial protection issues from the public and private sectors to discuss a range of

topics, including creating improved cybersecurity practices and strengthening cyber threat information sharing. The summit will also serve as the next step in the President's BuySecure Initiative, will help advance national efforts the government has led on consumer financial protection and critical infrastructure cybersecurity, and will build on efforts to improve cybersecurity at a wide range of companies.

5. Data as a Public Resource

The report urged agencies across government to consider data as a national, public resource, and make it broadly available to the public wherever possible. This effort continues the Obama Administration's commitment to open data and open government from the first day of this Administration. To date, there are over 134,000 datasets available on Data.gov for public use. The Administration has made great strides towards bringing technologists into government through the creation of the United States Digital Service, 18F, and the Presidential Innovation Fellowship to ensure that the government continues to meet the needs of Americans who expect the high quality digital content, as well as make data open and usable to the public.

Continue Making Government Data Available to the Public

The Administration has launched a series of Open Data Initiatives that have unleashed large volumes of valuable data in areas such as health, energy, education, public safety, finance, and global development. For example, the Climate Data Initiative, launched in March 2014, leverages open climate data to fuel innovation and private sector entrepreneurship to advance climate change preparedness and community resilience through the development of data products, tools, and applications that are geared toward solving real-life challenges.

This Administration is committed to making open and machine-readable data the default for government information. Federal agencies have continued to increase the quantity and quality of open data over the past year. Each quarter, federal agencies add additional datasets to their Public Data Listings. Data.gov automatically updates its inventory by harvesting the Public Data Listings each day. Nearly every agency has data listed on Data.gov.

Adopting Open Data Best Practices

Many federal agencies have adopted new open data processes to better manage their data at an organizational level. For example, over the past year, NASA has continued to develop an agency-level NASA Information Architecture Management (NIAM) process to share and reuse data from across agency components. Through the NIAM process, NASA significantly improved its common metadata, contract language, and search capacity, and as a result, NASA increased its Enterprise Data Inventory from 25 datasets to more than 3,800 datasets between November 2013 and November 2014.

Increased customer engagement is helping to improve the federal open data policy. For example, agencies have learned that one of the most common complaints of data users is the use of PDF—rather than machine-readable—formats. OMB and OSTP are now working with agencies to reduce the number of PDFs and make machine-readability the standard for all government data.

Issues Needing Further Attention

Some efforts await Congressional or stakeholder action. For instance, efforts on Capitol Hill to amend the almost 30-year-old Electronic Communications Privacy Act have seen little progress since the report was issued. The report recommended that Congress amend ECPA to ensure the standard of protection for online, digital content is consistent with that afforded in the physical world—including by removing archaic distinctions between email left unread or over a certain age.

In 2012, the Administration expressed support for the multistakeholder development of a Do Not Track standard that could be used by consumers regardless of browser preference or operating system. This was a novel multi-stakeholder effort, bringing together the technical community, advertisers, publishers, and privacy experts, and the big data and privacy working group called for the initiative to continue its efforts. Disappointingly—and despite no downturn in consumer interest—there have been delays in moving this initiative forward. Stakeholders should recommit to developing new voluntary tools, including Do Not Track, to safeguard users' privacy.

Conclusion

Less than a year after the release of the big data and privacy working group's findings, the Obama Administration has made significant progress in furthering the majority of the recommendations made in the big data and privacy report. Policy development remains actively underway on complex recommendations, including extending more privacy protections to non-U.S. persons and scaling best practices in data management across government agencies. And in big data and discrimination, the civil rights and privacy communities will continue to play an active and critical role in driving the conversation, partnering with the federal government, and surfacing new issues for consideration in this new field.

Beyond the conclusions of the big data and privacy working group, the insights in the report have also had influence on Administration policy. In his State of the Union address, President Obama announced an ambitious plan to advance understanding of precision medicine, an emerging field that holds the promise of revolutionizing how we improve health and treat disease. Leveraging advances in genomics, clinical practice, big data technology, and other fields, the Precision Medicine Initiative will seek to create a one-million-strong national research cohort and to accelerate discovery of tailored treatments for cancers. Data security and patient privacy will be paramount to the Precision Medicine Initiative. The effort will incorporate the lessons learned by other federal agencies and the issues identified in the big data and privacy report and solicit input from a diverse range of privacy stakeholders from the earliest days in order to integrate rigorous privacy protections throughout the program.

The big data and privacy working group concluded that, despite the newness of the field, big data is already saving lives, making the economy and the government work better, and saving taxpayer dollars along the way. Big data will continue to contribute to and shape our society, and the Obama Administration will continue working to ensure that government and civil society strive to harness the power of these technologies while protecting privacy and preventing harmful outcomes.