March 31, 2014

Via e-mail: bigdata@ostp.gov

Nicole Wong, Esq.
Big Data Study
Office of Science and Technology Policy
Eisenhower Executive Office Building
1650 Pennsylvania Avenue NW
Washington, DC 20502

Re: Public Comments, Big Data RFI

Dear Ms. Wong:

The Future of Privacy Forum (FPF) is a think tank seeking to advance responsible data practices and is supported by leaders in business, academia, and consumer advocacy.[1] FPF submits these Comments in response to the White House Office of Science and Technology Policy (OSTP) Request for Information (RFI) dated March 4, 2014. In the RFI, the OSTP seeks public comment on how best to ensure innovation and maximize the opportunities and free flow of big data while minimizing any risks to privacy.[2]

Unlocking the value of data and instituting responsible data practices go hand-in-hand, and both have been an important focus of FPF's work since our founding in 2008. FPF recognizes the enormous potential benefits to consumers and to society from sophisticated data analytics,[3] yet FPF also understands that taking advantage of big data may require evolving how we implement traditional privacy principles. Through our work on inter-connected devices and applications and the emerging Internet of Things, FPF has acquired experience with the technologies involved in data collection and use. FPF appreciates this opportunity to provide Comments and share its insights into how best to promote the benefits of big data while minimizing any resulting privacy risks or harms.

Responding to the President's call to review how big data is changing our society, OSTP's Big Data Review has been a helpful exercise in soliciting thought leadership from academics, researchers, and industry. There is much that can be done to promote innovation in a way that

[1] The views herein do not necessarily reflect those of the Advisory Board or supporters of the Future of Privacy Forum.
[2] White House Office of Science and Technology Policy, Government "Big Data": Request for Information, 79 Fed. Reg. 12,251 (Mar. 4, 2014).
[3] Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 243-51 (2013).

advances privacy, and we are pleased to provide our recommendations. Specifically, we recommend that the OSTP Big Data Review report:

1) **Embrace a flexible application of Fair Information Practice Principles (FIPPs).** Traditional FIPPs have guided privacy policy nationally and around the globe for more than 40 years, and the White House Consumer Privacy Bill of Rights is the most recent effort to carry these principles forward into a world of big data. FPF supports the continued reliance on the FIPPs and believes they remain flexible enough to address many of the challenges posed by big data when applied in a practical, use-based manner. Our Comments recommend a nuanced approach to their applicability that accounts for modern day technical realities.

2) **Promote the benefits of big data in society.** Researchers, academics, and industry have demonstrated how big data can be useful in driving economic growth, advancing public safety and health, and improving our schools. Yet, privacy advocates and the public appear skeptical of these benefits in the face of certain outlier uses. More work is needed to understand the ways big data is already improving society and making businesses more efficient and innovative. This report should highlight the importance of big data's benefits and identify additional opportunities to promote positive uses of big data.

3) **Support efforts to advance practical de-identification, including policy and technological solutions.** While the Federal Trade Commission (FTC) has acknowledged that data that is effectively de-identified poses no significant privacy risk, there remains considerable debate over what effective de-identification requires. FPF believes that technical anonymization measures are only one component of effective de-identification. Instead, a broader understanding that takes into account how administrative and legal safeguards, as well as whether data is public or non-public, should inform conversations about effective de-identification procedures.

4) **Encourage additional work to frame context and promote enhanced transparency.** The context in which data is collected and used is an important part of understanding individuals' expectations, and context is a key principle in both the Consumer Privacy Bill of Rights and the FTC Privacy Framework. Respect for context is an increasingly important privacy principle, yet more work by academics, industry, and policymakers is needed about how to properly frame and define this principle. The Department of Commerce-led Internet Policy Task Force (IPTF) should continue its work convening stakeholders and hold programs that could help frame context in an age of big data. At the same time, another important tool that can be used to promote public trust in big data is enhanced transparency efforts. In particular, FPF has called for more transparency surrounding high-level decisional criteria that organizations may use to make decisions about individuals.

5) **Encourage efforts to promote accountability by organizations working with big data.** Data privacy frameworks increasingly rely on organizational accountability to ensure responsible data stewardship. In the context of big data, FPF supports the further development of the concept of internal review boards that could help companies weigh the benefits and risks of data uses. In conjunction with the evolving role of the privacy professional, accountability measures can be put in place to ensure big data projects take privacy considerations into account.

6) **Promote government leadership on big data through its own procedures and practices.** The federal government is one of the largest producers and users of data, and, as a result, the government may inform industry practice and help demonstrate the value of data through its own uses of big data across and among agencies. The Federal Chief Information Officer (CIO) Council is particularly well-positioned to ensure the federal government can maximize the potential of big data with an eye toward privacy protection.

7) **Promote global efforts to facilitate interoperability.** Recent privacy developments in the Asia Pacific and the European Union have given new life to constructive collaboration on the cross jurisdictional issues presented by big data. FPF urges government to actively promote and maintain existing frameworks to facilitate interoperability, including the US-EU Safe Harbor and the Asia Pacific Economic Cooperation's (APEC) Cross Border Privacy Rules (CBPR) System**.**

These broad next steps are suggested as a helpful beginning to the work that needs to be done. In the remainder of this submission, we respond to the questions posed in the RFI.

(1) What are the public policy implications of the collection, storage, analysis, and use of big data?

Big data may be one of the biggest public policy challenges of our time.[4] The debate surrounding big data will ask policy makers to pit compelling interests such as national security, public health and safety, and sustainable development against risks to personal autonomy from high-tech profiling and discrimination, increasingly-automated decision making, inaccuracies and opacity in data analysis, and strains in traditional legal protections.[5] However, the traditional Fair Information Practice Principles (FIPPs) remain flexible enough to address many of these concerns when applied in a practical, use-based manner. What is needed is additional research on the benefits of big data and on how to advance practical de-identification and other measures to protect privacy.

I. Big Data and the Fair Information Practice Principles

There is considerable dispute today over how best to properly calibrate the FIPPs to protect privacy and encourage innovative uses of data. On one hand, some increasingly suggest that foundational privacy practices such as a notice and choice and purpose limitation are either impractical or less relevant due to big data and other emerging technologies.[6] While privacy advocates and regulators recognize limitations with our notice and choice framework, they worry that big data may provide an excuse to override individual rights in order to facilitate intrusive marketing or ubiquitous surveillance. FPF would caution against disposing of sound principles that have guided privacy policy for more than forty years. Our Comments advocate for a nuanced approach, based upon a practical application of the FIPPs that accounts for modern day technical realities around collection and use of personal data.

[4] Jules Polonetsky, Omer Tene & Christopher Wolf, *How To Solve the President's Big Data Challenge*, IAPP PRIVACY PERSPECTIVES (Jan. 31, 2014), https://www.privacyassociation.org/privacy_perspectives/post/how_to_solve_the_presidents_big_data_challenge.
[5] Civil Rights Principles for the Era of Big Data, http://www.civilrights.org/press/2014/civil-rights-principles-big-data.html (last visited March 15, 2013).
[6] For example, the growing network of smart, connected devices known as the "Internet of Things" is commonly understood to rely upon the capture, sharing, and use of data, including data about who we are and what we do at any given moment. *See, e.g.*, Bill Wasik, *Welcome to the Programmable World*, WIRED (May 14, 2013), http://www.wired.com/gadgetlab/2013/05/internet-of-things/.

In their various formulations, the FIPPs establish core principles guiding the collection, use, and disclosure of data.[7] Some of the most important FIPPs are: (1) Notice – individuals should be provided with timely notice of how their data will be collected, used, and disclosed; (2) Choice – individuals should be given choices about whether and how their data will be used; (3) Purpose Specification – the purposes for which personal data are collected should be specified prior to or at the time of collection; and (4) Use Limitation – personal data should only be used for those purposes specified prior to or at the time of collection; and (5) Data Minimization – organizations should seek to limit the amount of personal data they collect and that might be retained.[8] These principles are each challenged by big data in different ways.

The White House Consumer Privacy Bill of Rights has recognized this challenge. Based on the FIPPs, the general principles put forward by the Administration's privacy framework explicitly afford companies discretion in how they implement them. This flexibility was designed both to promote innovation and to "encourage effective privacy protections by allowing companies, informed by input from consumers and other stakeholders, to address the privacy issues that are likely to be most important to their customers and users, rather than requiring companies to adhere to a single, rigid set of requirements."[9]

A. Notice and Choice: The Need for Flexibility

Notice is often considered the most "fundamental" principle of privacy protection.[10] Yet there is wide acknowledgement that a privacy framework based on notice and choice has significant limitations.[11] The vast majority of consumers do not read privacy policies,[12] and further, studies have shown that consumers make privacy decisions not based on policies but rather on the context in which they are presented by a use of their data.[13] In the age of big data, the implementation of notice and choice through detailed privacy policies may only result in the publication of even more unread policies. Furthermore, notice and choice presents particular problems for connected devices or other "smart" technologies that will not be equipped with interactive screens or other easily accessible user interfaces. Information collected in "public" spaces and used for data analytics may also prove problematic.

---

[7] The FIPPs generally have been thought of as establishing high-level guidelines for promoting privacy. They do not establish specific rules prescribing how organizations must protect privacy in all contexts, but rather they provide principles that can inform the implementation of specific codes of practice. Initially proposed in a 1973 advisory committee report for the Department of Housing, Education, and Welfare, the FIPPs emerged due to concern about the increased use of personal data in record-keeping systems. Subsequently, the FIPPs have become the basis of global privacy law and remain relevant in a world of big data.

[8] *See* Christopher Wolf & Jules Polonetsky, *An Updated Privacy Paradigm for the "Internet of Things"* (2013), http://www.futureofprivacy.org/wp-content/uploads/Wolf-and-Polonetsky-An-Updated-Privacy-Paradigm-for-the-"Internet-of-Things"-11-19-2013.pdf.

[9] WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY 2 (2012), http://www.whitehouse.gov/sites/default/files/privacy-final.pdf [hereinafter WHITE HOUSE BLUEPRINT].

[10] FEDERAL TRADE COMMISSION, PRIVACY ONLINE: A REPORT TO CONGRESS 7 (1998).

[11] Fred Cate, *Looking Beyond Notice and Choice*, PRIVACY & SECURITY LAW REPORT (Mar. 29, 2010), http://www.hunton.com/files/Publication/f69663d7-4348-4dac-b448-3b6c4687345e/Presentation/PublicationAttachment/dfdad615-e631-49c6-9499-ead6c2ada0c5/Looking_Beyond_Notice_and_Choice_3.10.pdf (citing Former FTC Chairman Jon Liebowitz conceding that the "notice and choice" regime offered by the FIPPs hasn't "worked quite as well as we would like.").

[12] Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J. L. & POL'Y FOR INFO. SOC'Y 543 (2008).

[13] *See, e.g.*, Alessandro Acquisti et al., What Is Privacy Worth? 27-28 (2010) (unpublished manuscript), *available at* http://www.heinz.cmu.edu/~acquisti/papers/acquisti-ISR-worth.pdf.

Although technological solutions may help to facilitate notice and choice options, it will be impractical to premise data collection and use in the world of big data and other emerging technologies based solely on traditional implementations of notice and choice. For that matter, a number of data applications may not require any integration of privacy protections. Consider a smart TV that learns the volume preferences of particular users and adjusts its volume accordingly – if that information is not transmitted out of the TV it should raise few issues. Most machine-to-machine communications of contextually aware devices also make notice and choice unnecessary, especially if any information flows for these devices are contained.

That said, privacy policies still have value. They remain helpful as accountability and enforcement mechanisms: they set the boundaries for data use by businesses beyond those that might be prescribed by law and they create enforceable legal obligations. Disclosure requirements by themselves can force companies to evaluate their privacy practices and instill discipline in how they treat consumer information.[14]

Flexibility will be especially important with respect to the concepts of notice and choice. There remains much uncertainty over what information matters for disclosure, and this problem is only exacerbated by big data. Organizations need to think creatively about how to provide consumers with meaningful insight into commercial data practices, and regulators and policymakers should encourage these efforts.[15]

This challenge, however, is recognized by the Consumer Privacy Bill of Rights, which recommends organizations seek innovative ways to provide consumers with more individual control, and if that remains impractical, organizations should embrace and augment other FIPPs or elements of the Consumer Privacy Bill of Rights in order to adequately protect consumer privacy.[16] The OSTP Big Data Review report should call for renewed efforts by industry to develop new models to inform individuals about the collection and use of their personal data. Techniques to inform consumers of data practices might include symbols, short phrases, colors, diagrams, or any of the tools otherwise available to designers seeking to provide users with an engaging user experience. Engaging consumers about data use should be viewed as an essential feature and a core part of the user experience. In the end, design features that "communicate" information to users may be more helpful than traditional notice models.

B. Purpose Specification & Use Limitation: Context Is Key

One of the exciting challenges presented by big data is that much of the new value from data is being discovered in surprising ways.[17] Consider the innovations pioneered by the United Nations Global Pulse that are enabled by the analysis of mobile phone data. Global Pulse has helped us understand mobility, social interaction and economic activity.[18] By analyzing mobile interactions,

---

[14] *See* Peter P. Swire, *The Surprising Virtues of the New Financial Privacy Law*, 86 MINN. L. REV. 1263, 1314 (2002).

[15] Testimony Before the California State Assembly Joint Committee Hearing on Privacy (Dec. 12, 2103) (statement of Jules Polonetsky, Executive Director, Future of Privacy Forum, at 3), *available at* http://www.futureofprivacy.org/wp-content/uploads/CA-Assembly-Hearing-Privacy-Policies_Does-Disclosure-Transparency-Adequately-Protect-Consumers-Privacy-Final.pdf (citing Susanna Kim Ripken, *The Dangers and Drawbacks of the Disclosure Antidote: Toward a More Substantive Approach to Securities Regulation*, 58 BAYLOR L. REV. 139, 147 (2006) (calling for regulators to "lay aside the gospel of disclosure in favor of more substantive laws that regulate conduct directly")).

[16] WHITE HOUSE BLUEPRINT, *supra* note 9, at 13.

[17] *E.g.*, VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (2013).

[18] Robert Kirkpatrick, Beyond Targeted Ads: Big Data for a Better World (2012), *available at*

UN researchers were able to examine the post-earthquake population migration caused by the Haiti earthquake. Global Pulse has been able to track the spread of disease and better understand socio-economic activity in a number of countries around the world. However, under our traditional privacy frameworks, some valuable uses of data may be constrained.

Most privacy regimes endorse a principle of use limitation, which is generally implemented by requiring that personal information be used *only* as specified at the time of collection.[19] Most of the innovative secondary uses of information – including breakthroughs in medicine, data security, or energy usage – are impossible to anticipate when notice is first provided, often long before a new benefit is uncovered through data analysis.[20] Companies can neither provide notice for a purpose that is yet to exist, nor can consumers provide informed consent for an unknown.[21]

However, these principles may be implemented by instead limiting the use of information based upon the *context* in which it is collected.[22] Often, context is understood to mean that personal information should be used only in ways that individuals would expect given the context in which information was disclosed and collected. However, there are uses of data that may be outside individual expectations but have high societal value and minimal privacy impact that should be encouraged. More work is needed to define and frame context.

C. Data Minimization: Moving Toward More Accountability Measures

While it has been overshadowed by the principles of notice and choice,[23] data minimization has long been another important traditional privacy practice.[24] Data minimization promotes privacy by limiting the amount of personal information in circulation.[25] Yet it is not clear that minimizing information collection is always a practical approach to privacy in the age of big data.[26] Almost by definition, "big" data requires a significant amount of data to be available in order to discern previously unnoticed patterns and trends. As the Consumer Privacy Bill of Rights notes, "wide-ranging data collection may be essential for some familiar and socially beneficial internet services and applications."[27] These uses, as well as many others yet to be developed, would be stymied if companies were required to limit the amount of data they collect.

---

http://www.slideshare.net/unglobalpulse/strata-14934034.

[19] *See, e.g.*, European Parliament and Council Directive 95/46/EC - on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995 O.J. (L 281) 31, *available at* http://www.refworld.org/docid/3ddcc1c74.html (last visited Mar. 15, 2014).

[20] SCHÖNBERGER & CUKIER, *supra* note 17, at 153.

[21] *Id.*

[22] Wolf & Polonetsky, *supra* note 8, at 9.

[23] Fred Cate, *The Failure of the Fair Information Practice Principles* 15 (2009), http://www.informationpolicycentre.com/files/Uploads/Documents/Centre/Failure_of_Fair_Information_Practice_Principles.pdf

[24] *See* OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, ORG. FOR ECON. CO-OPERATION & DEV. (Sept. 23, 1980), http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html. Data minimization involves limiting an organization's collection of personal data to the minimum extent necessary to obtain specified and legitimate goals. The principle further instructs organizations to delete data that is no longer used for the purposes for which it was originally collected, and to implement restrictive policies with respect to the retention of personal data in identifiable form.

[25] With less data to process and analyze, many believe that companies will have less capability to use data in new, privacy-invasive ways – and consumers will be protected from unwarranted access to their information. *See, e.g.*, Justin Brookman & G.S. Hans, *Why Collection Matters* (2013), http://www.futureofprivacy.org/wp-content/uploads/Brookman-Why-Collection-Matters.pdf.

[26] Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63 (2012), http://www.stanfordlawreview.org/online/privacy-paradox/big-data.

[27] WHITE HOUSE BLUEPRINT, *supra* note 9, at 21.

There is still a role for sensible retention policies and efforts to reasonably limit data collection should not be dismissed out-of-hand. However, concerns around data collection and use may be mitigated through additional accountability measures, such as internal controls and internal review boards, which we discuss below. Further, when organizations use adequately de-identified data sets, their use of that data mitigates privacy risk, which demonstrates how further research around de-identification could prove helpful within the context of big data. Thus, a more sophisticated analysis of data minimization should take into account the de-identification and other privacy safeguards that have been implemented.

II. Advancing Practical De-Identification Solutions

Clarifying the scope of information subject to privacy law has become an increasingly important policy question. During this review's "Advancing the State of the Art in Technology and Practice" workshop at MIT, the question of what information is properly de-identified or anonymous emerged throughout the day's discussion.[28] Personally identifiable information (PII) is one of the central concepts in information privacy regulation, but there is no uniform definition of PII.[29] Similarly, there is no standard for what constitutes adequate de-identification of PII.[30]

This is important because resolving the spectrum of PII and non-PII also addresses some of the concerns facing traditional FIPPs. As the FTC acknowledged in its March 2012 report, *Protecting Consumer Privacy in an Era of Rapid Change*, data that has been effectively de-identified does not raise significant privacy concerns.[31] However, laws often turn on whether or not information is PII or not, and this bi-polar approach based on labeling information either "personally identifiable" or not is not appropriate given the messiness of big data.[32] FPF proposes that PII instead be defined based on a risk matrix taking into account the risk, intent, and potential consequences of re-identification, as opposed to a dichotomy between "identifiable" and "non-identifiable" data.

De-identification should be understood as a process that takes into account legal and administrative safeguards, as well as technical measures, in order to protect privacy. Unfortunately, at the moment, much of our discourse around de-identification focuses on the technical possibility of re-identification and the assumption that all data will be made publicly available.[33] While computer scientists have repeatedly shown that anonymized data, either released publicly or poorly de-identified, can be re-identified, organizations and policymakers must recognize that non-public data presents a lessened privacy risk than information released publicly.

---

[28] Big Data Privacy Workshop: Advancing the State of the Art in Technology and Practice, http://web.mit.edu/bigdata-priv/ (last visited Mar. 15, 2014).

[29] Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 NYU L. REV. 1814 (2011).

[30] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010) (arguing that "scientists have demonstrated that they can often 'reidentify' or 'deanonymize' individuals hidden in anonymized data with astonishing ease." *But see* Daniel Barth-Jones, *Re-Identification Risks and Myths, Superusers and Super Stories*, CONCURRING OPINIONS (Sept. 6, 2012), http://www.concurringopinions.com/archives/2012/09/re-identification-risks-and-myths-superusers-and-super-stories-part-ii-superusers-and-super-stories.html (citing Ohm's suggestion that public policy should not "inappropriately conflat[e] the rare and anecdotal accomplishments of notorious hackers with the actions of typical users . . .")).

[31] FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS 20-22 (2012), http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf [hereinafter FTC PRIVACY REPORT].

[32] *See generally* SCHÖNBERGER & CUKIER, *supra* note 17, at 32-49 (suggesting the value of big data may require an approach to data analysis that is "comfortable with disorder and uncertainty.").

[33] Yianni Lagos & Jules Polonetsky, *Public vs. Nonpublic Data: The Benefits of Administrative Control*, 66 STAN. L. REV. ONLINE 103 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/public-vs-nonpublic-data.

While any analysis of effective de-identification should consider the legal and administrative controls around data, there remains work to be done to advance technical de-identification measures. The Administration should support this effort through funding and by convening workshops that can help further de-identification research. This work could focus on de-identification techniques that maintain data utility for researchers and industry, and should help frame a de-identification debate that recognizes the value of non-technical safeguards.

III. Framing Context and Meaningful Transparency

A principle of respect for context relies on what individuals expect from their relationship with an organization. Consumers expect that companies will share their personal information with other companies to fulfill orders and that companies will use personal information to engage in first-party marketing.[34] When personal information is used in those ways or in others that individuals would reasonably expect, there is no privacy violation.

But respect for context can become difficult to meet when faced with innovative data practices.[35] Focusing solely on individual expectations not only hampers some benefits that could accrue to those individuals, but it also ignores that company-to-consumer relationships evolve. As the Consumer Privacy Bill of Rights recognizes, respect for context must admit that a relationship between an organization and an individual may change over time in ways not foreseeable at the time of collection, and that "such adaptive uses of personal data may be the source of innovations that benefit consumers."[36] Consider a company that collects personal fitness information from wearable sensors that track sleep, steps taken, pulse or weight. Analysis of such data, collected originally only to report basic details back to users, may yield unanticipated health insights that could be provided individually to users or used in the aggregate to advance medical knowledge. Rigidly and narrowly specifying context could trap knowledge that is available and critical to progress.

The challenge facing organizations and policymakers is that respect for context requires an appreciation for dynamic social and cultural norms.[37] Context includes not only an objective component, but also a number of subjective variables including an individual's level of trust and his perceived value from the use of his information.[38] Public-facing efforts to inform consumers about big data will be essential to provide individuals with more context around data practices. Companies could frame relationships by "setting the tone" for new products or novel uses of information. Even where new uses of data are contextually similar to existing uses, information and education are essential. Amazon serves as a prime example of this approach: its website is able to pursue a high degree of customization without violating consumer expectations, given its clear messaging about customization and its provision of a user interface frames how data is used.

---

[34] *See* WHITE HOUSE BLUEPRINT, *supra* note 9, at 16-17.
[35] Jules Polonetsky & Omer Tene, *It's Not How Much Data You Have, But How You Use It* 5 (2012), http://www.futureofprivacy.org/wp-content/uploads/FPF-White-Paper-Its-Not-How-Much-Data-You-Have-But-How-You-Use-It_FINAL1.pdf.
[36] WHITE HOUSE BLUEPRINT, *supra* note 9, at 16.
[37] Carolyn Nguyen, Director, Microsoft Technology Policy Group, Contextual Privacy, Address at the FTC Internet of Things Workshop (Nov. 19, 2013) (transcript available at http://www.ftc.gov/sites/default/files/documents/public_events/internet-things-privacy-security-connected-world/final_transcript.pdf).
[38] *Id.*

To complement a principle focused on respect for context, organizations must be much more transparent about how they are using data. Many of the concerns around big data applications center on worries about untoward data usage, and enhanced transparency may help alleviate fears that an individual's personal information is somehow being used against them. Transparency can be a tool that can help demystify big data.[39] For example, organizations should disclose the criteria underlying their decision-making processes to the extent possible without compromising their trade secrets or intellectual property rights. While there are practical difficulties in requiring these disclosures, distinctions can be drawn between sensitive, proprietary algorithms and high-level decisional criteria.

Further transparency efforts have been endorsed by both the Consumer Privacy Bill of Rights and the FTC's 2012 privacy report.[40] But transparency cannot simply mean better privacy policies. Instead, policymakers should encourage companies to engage with consumers in a meaningful conversation where both parties' interests and expectations can be aligned.[41] FPF has previously called for the "featurization" of data, transforming data analysis into a consumer-side application by granting individual access to their personal data in intelligible, machine-readable forms. Mechanisms such as personal clouds or data stores will allow individuals to contract with third-parties who would get permission to selectively access certain categories of their data to provide further analysis or value-added services.[42] "Featurization" will allow individuals to declare their own policies, preferences and terms of engagement, and do it in ways that can be automated both for them and for the companies they engage.[43]

IV. Enhanced Accountability through Internal Review Boards

Big data may warrant a shift in focus toward accountability mechanisms that ensure organizations are responsibly managing personal information.[44] Several privacy scholars have suggested that our current privacy framework stresses mere compliance, when emphasizing institutional accountability may be more necessary to promote better data stewardship.[45] While there are many strategies to augment accountability in the age of big data, it will be important for organizations to engage in a practical balancing of privacy considerations and data use.

A formalized review mechanism could help to review and approve innovative data projects.[46] Some have also called for big data "algorithmists" that could evaluate the selection of data sources, the choice of analytical tools, and the interpretation of any predictive results.[47] As organizations

---

[39] Tene & Polonetsky, *supra* note 3, at 270-72.
[40] WHITE HOUSE BLUEPRINT, *supra* note 9, at 16-17; FTC PRIVACY REPORT, *supra* note 31, at 60.
[41] *See, e.g.*, Omer Tene & Jules Polonetsky, *A Theory of Creepy: Technology, Privacy and Shifting Social Norms* , 16 YALE J.L. & TECH. 59 (2013).
[42] Tene & Polonetsky, *supra* note 3, at 263-70.
Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 243-51 (2013).
[43] The rise of privacy and reputation management vendors points to a future where organizations will be able to unlock additional value by collaborating with their users. One interesting first step is the launch of "About the Data" by Acxiom, the nation's largest data broker. "About the Data" is a consumer-facing tool that gives individuals control over certain categories of information (such as personal characteristics, interests, and finances) gathered by Acxiom. The site allows consumers to correct information, suppress any data they see, or opt-out of Acxiom's marketing profile system altogether via an approachable user interface.
[44] *See* Fred H. Cate & Viktor Mayer-Schoenberger, Data Use and Impact Global Workshop (2013), http://cacr.iu.edu/sites/cacr.iu.edu/files/Use_Workshop_Report.pdf.
[45] *Id.* at 5.
[46] Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 STAN. L. REV. ONLINE 97 (2013).
[47] SCHÖNBERGER & CUKIER, *supra* note 17, at 180.

increasingly face interesting new proposals for using data, these professionals could operate across the public and private sectors and conduct cost-benefit analyses of data uses.

Industry increasingly faces ethical considerations over how to minimize data risks while maximizing benefits to all parties. Formal review processes may serve as an effective tool to infuse ethical considerations into data analysis. Institutional review boards (IRBs) were the chief regulatory response to decades of questionable ethical decisions in the field of human subject testing; big data internal review boards could similarly serve as a proactive response to concerns regarding data misuse. In many respects, these review boards would be a further expansion of the role of privacy professionals within the industry today. While creating internal review boards would present a unique set of challenges, encouraging companies to create sophisticated structures and personnel to grapple with these issues and provide oversight would be invaluable.

Any successful approach to big data must be guided by a cost-benefit analysis that takes into account exactly how the benefits of big data will be distributed. So far, our procedural frameworks are largely focused on traditional privacy risks and assessing what measures can be taken to mitigate those risks. In 2010, for example, the Department of Commerce's Internet Policy Task Force endorsed the use of privacy impact assessments (PIAs) both to help organizations decide whether it is appropriate to engage in innovative data uses and to identify alternative approaches that could reduce relevant privacy risks.[48] However, human research IRBs also take into account anticipated benefits and even the importance of any knowledge that may result from research.[49]

FPF is exploring a framework to provide a similar accounting of the rewards of big data. While organizations and privacy professionals have developed expertise at evaluating risk, we believe decision makers need better processes to evaluate how to assess, prioritize, and to the extent possible, quantify a project's potential benefits. To that end, we intend to propose a methodology that assesses a project's value based upon several criteria, including culture-specific values and probability of success. FPF is eager to engage in further discussion about how best to develop big data review mechanisms.

(2) What types of uses of big data could measurably improve outcomes or productivity with further government action, funding, or research? What types of uses of big data raise the most public policy concerns? Are there specific sectors or types of uses that should receive more government and/or public attention?

One of the chief challenges facing big data analytics is determining how much of "big data" is new, or whether it is simply another buzzword. There have been detailed efforts to show how the emerging technologies fueled by big data are reshaping society.[50]

I. The Social Ramifications of Predictive Analytics

The fundamental problem posed by big data may be less a question of how it impacts our privacy and more that it upsets our society's sense of fairness. The debate around big data is often couched

[48] U.S. DEP'T OF COMMERCE, INTERNET POL'Y TASK FORCE, COMMERCIAL DATA PRIVACY AND INNOVATION IN THE INTERNET ECONOMY: A DYNAMIC POLICY FRAMEWORK 34-35 (2010), http://www.commerce.gov/sites/default/files/documents/2010/december/iptf-privacy-green-paper.pdf.
[49] 45 CFR § 46.111.
[50] *E.g.*, RICK SMOLAN & JENNIFER ERWITT, THE HUMAN FACE OF BIG DATA (2012).

as something that implicates traditional privacy principles and that the use and inferences drawn from our data invade our privacy, but this obscures the larger public policy challenge. The real concerns presented by big data are increasingly abstract or inchoate risks that may have very little to do with privacy practices per se.

A. Shifting User Expectations: "Creepiness"

Since the revelation that Target was able to predict a teenager's pregnancy before her family was even aware of it, much of the concern surrounding predictive analytics is that it is somehow "creepy." The Target example is illustrative precisely because it did not involve any explicit breach of the FIPPs, and was not explicitly harmful to the consumer in question. Instead, it was a novel use of data that dramatically upset individual expectations. Disrupting one's expectations can lead to unexpected benefits, but it also limits an individual's ability to feel comfortable or in control.[51]

While creepiness is inherently subjective, creepy behaviors are detrimental to the development of any trust-based relationship – whether between friends, consumer and company, or government and citizen.[52] Due to a lack of trust, individuals have been quick to dismiss the benefits of big data as a result of some of its more surprising results. The use of big data is outpacing the evolution of social norms, and addressing creepiness may simply be a matter of organizations exploring how to set the right tone when seeking intimate relationships with individuals via their data.[53] Self-regulatory frameworks are likely to be the best mechanism to address shifting cultural norms, and industry should be encouraged to be proactive in this regard.

B. Potential Concerns Surrounding Civil Liberties

There are other broad concerns about big data that are outside the ambit of traditional privacy law. Recently, critics, including some of the United States' leading civil rights organizations, have argued that big data could be the "civil rights" issue of this generation.[54] In particular, there are worries that big data undermines equal opportunity and equal justice through hidden or new forms of discrimination.[55] Big data could achieve these harms by contributing to currently illegal practices, allowing otherwise unlawful activity to go undetected due to a lack of transparency or access surrounding data analysis.[56] Alternatively, big data may introduce societal biases that may impact protected classes or otherwise vulnerable populations disproportionately or unfairly.[57]

The United States has enacted a series of powerful legislative remedies to combat discrimination in the context of employment, education, housing, and credit worthiness. These laws specifically

---

[51] Francis T. McAndrew & Sara S. Koehnke, (On the Nature of) Creepiness, Poster presented at the annual meeting of the Society for Personality and Social Psychology (SPSP), Jan 18, 2013, *available at* http://www.academia.edu/2465121/Creepiness ("Being 'creeped out' is an evolved adaptive emotional response to ambiguity about the presence of threat that enables us to maintain vigilance during times of uncertainty.").

[52] *See generally id.*

[53] *See* Omer Tene & Jules Polonetsky, *A Theory of Creepy: Technology, Privacy and Shifting Social Norms* , 16 YALE J.L. & TECH. 59 (2013).

[54] Alistair Croll, *Big Data Is Our Generation's Civil Rights Issue, and We Don't Know It*, SOLVE FOR INTERESTING (July 31, 2012), http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it/; Civil Rights Principles for the Age of Big Data, *supra* note 5.

[55] Civil Rights Principles for the Age of Big Data, *supra* note 5.

[56] Pam Dixon, *On Making Consumer Scoring More Fair and Transparent*, IAPP PRIVACY PERSPECTIVES (Mar. 19, 2014), https://www.privacyassociation.org/privacy_perspectives/post/on_making_consumer_scoring_more_fair_and_transparent.

[57] See Kate Crawford, *The Hidden Biases in Big Data*, HBR BLOG NETWORK (Apr. 1, 2013), http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/.

protect access to certain opportunities by prohibiting organizations from taking into account certain factors. For example, Title VII of the Civil Rights Act of 1964 prohibits employers from discriminating against applicants and employees on the basis of race, color, religion, sex, and national origin.[58] The Equal Credit Opportunity Act forbids creditors from asking about a candidate's marital status or plans to have children.[59] Our antidiscrimination laws have even had to take new technologies into account: the Genetic Information Nondiscrimination Act of 2008, for example, prohibits employers from using an applicant's or an employee's genetic information as the basis of an employment decision, and it also limits the ability of health insurance organizations to deny coverage based solely on a genetic predisposition to develop a disease.[60]

Antidiscrimination laws are not truly data privacy laws, however. Instead, they work to address classifications and decisions that society has deemed either irrelevant or illegitimate. Big data makes it easier to discover an individual's race, religion, or gender, and it also encourages ever more granular categorization and segmentation of individuals. The challenge is that there are no clear guidelines as to where value-added personalization and segmentation – which may provide positive consumer benefits in some cases – turn into harmful discrimination.[61] We already have legal protections in place that would prohibit the use of big data to engage in certain kinds of specific discrimination. Further academic and expert analysis is necessary in order to understand which of the claimed data practices are already illegal and merely need additional enforcement and which create new uses that warrant further policy analysis.

However, the bigger question is whether big data may impact individuals or classes of individuals in ways that we might deem unfair. Take the example of an Atlanta man who returned from his honeymoon to find his credit limit slashed from $10,800 to $3,800 because he had used his credit card at locations where *others* were likely to have a poor repayment history.[62] Is this an efficient use of data analysis or fundamentally unfair? Are there industry practices or remedies that could avoid problems and concerns? If new data practices are deemed unfair, illegitimate, or have a disparate impact against an already vulnerable group, more work is necessary to understand how best to address these practices. These issues are already receiving priority attention at the FTC, which in 2012 entered into an $800,000 settlement with Spokeo for marketing personal information to employers in violation of the Fair Credit Reporting Act,[63] and more recently, held a workshop on the use of predictive analytics to create consumer scores.[64] FPF and other groups could play a positive role by further exploring these issues.

C. Transparency and Opacity: Filter Bubbles and Surveillance

Another more inchoate concern presented by big data is that it may allow institutions to know more about an individual than that individual even knows. Even if organizations have the best of

[58] 42 U.S.C. § 2000e-2.
[59] 15 U.S.C. § 1691.
[60] Pub.L. 110–233, 122 Stat. 881
[61] Michael Schrage, *Big Data's Dangerous New Era of Discrimination*, HBR BLOG NETWORK (Jan. 29, 2014), http://blogs.hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination/.
[62] *See* Lori Andrews, *Facebook Is Using You*, N.Y. TIMES (Feb. 4, 2012), http://www.nytimes.com/2012/02/05/opinion/sunday/facebook-is-using-you.html.
[63] Press Release, Fed. Trade Comm'n., Spokeo to Pay $800,000 to Settle FTC Charges Company Allegedly Marketed Information to Employers and Recruiters in Violation of FCRA (June 12, 2012), http://www.ftc.gov/news-events/press-releases/2012/06/spokeo-pay-800000-settle-ftc-charges-company-allegedly-marketed.
[64] Fed. Trade Comm'n., Spring Privacy Series: Alternative Scoring Products (Mar. 19, 2014), http://www.ftc.gov/news-events/events-calendar/2014/03/spring-privacy-series-alternative-scoring-products.

intentions, the knowledge gained from analysis of big data can quickly lead to over-personalization. Individuals are more easily segmented, classified, and potentially placed into "filter bubbles" at the expense of their autonomy.[65] This also produces a more segregated society.

Feelings of being surveilled, which can arise from the continuous collection and use of our information, may also impact how people behavior, causing a chilling effect on civil discourse. For example, pervasive web tracking presents the possibility that people may avoid certain searches or sources of information out of fear that accessing that information would reveal interests, medical conditions, or other characteristics they would prefer be kept hidden.[66] Combined with a lack of transparency about how this information is being used, individuals may feel anxiety over consequential decisions about them being made opaquely, inducing a sense of powerlessness.[67]

The challenge is that some of the new risks associated with big data are not easily mapped to recognizable harms or are difficult to link to accepted privacy risks. To what degree they even present real challenges to society remains an open question. More work is needed by researchers to determine how these abstract concerns should inform any big data privacy analysis.

II. A Positive Role for Government

The White House has previously shown significant commitment to using big data to advance the national good, and it can do more to highlight these benefits and alleviate any concerns. Government can play a pivotal role as a convener, encouraging the benefits of the big data while promoting efforts within industry to advance the framework presented by the Consumer Privacy Bill of Rights. At the same time, it can also provide leadership and guidance on big data through its own procedures and practices.[68] The federal government is one of the biggest producers of big data. More than $200 million was committed in 2012 as part of a National Big Data Research and Development Initiative.[69] Last year, the Administration continued to be active in convening multiple stakeholders to explore big data applications that can improve economic growth, job creation, education, health, energy, sustainability, public safety, advanced manufacturing, science and engineering, and global development.[70]

The federal government, through Open Data efforts and day-to-day interactions with the public, generate massive amounts of data. In 2012, for example, the President launched an initiative entitled Digital Government: Building a 21st Century Platform to Better Serve the American People.[71] The aims of the initiative are to offer the public access to government information and

[65] *E.g.*, ELI PARISER, THE FILTER BUBBLE: HOW THE NEW PERSONALIZED WEB IS CHANGING WHAT WE READ AND HOW WE THINK (2012).
[66] Felix Wu, *Big Data Threats* 2 (2013), http://www.futureofprivacy.org/wp-content/uploads/Wu-Big-Data-Threats.pdf.
[67] *Id.*
[68] Adelaide O'Brien, Iron Mountain, *The Impact of Big Data on Government* (Oct. 2012), http://www.ironmountain.com/Knowledge-Center/Reference-Library/View-by-Document-Type/White-Papers-Briefs/Sponsored/IDC/The-Impact-of-Big-Data-on-Government.aspx.
[69] Press Release, White House Office of Science & Tech., Obama Administration Unveils "Big Data" Initiative: Announces $200 Million in New R&D Investments (Mar. 29, 2012), http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf.
[70] Big Data Senior Steering Group, The Networking and Information Technology Research and Development Program, http://www.nitrd.gov/nitrdgroups/index.php?title=Big_Data_(BD_SSG)#title (last visited Mar. 15, 2014).
[71] White House, Digital Government: Building a 21st Century Platform to Better Serve the American People, http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html (last visiting Mar. 15, 2014).

services anytime, anywhere, on any device, and to better leverage the rich wealth of federal data by promoting open data and machine-readable information.

Increasingly, government agencies are also using data in more creative ways. The Consumer Financial Protection Bureau (CFPB) has been especially active in using data analytics tools, arguing that "a 21st-century agency should use 21st-century tools."[72] In order to fulfill its mission and effectively monitor financial practices, the bureau has gathered vast amounts of consumer finance data on information varying from overdraft fees to credit card add-on products, and it is building databases that will integrate consumer credit information with loan and property records.[73] Yet, the CFPB also serves as an example of some the key privacy challenges that may come with innovative data use. The bureau has received criticism over concerns about its transparency and accountability when it comes to using individual financial data to police bank behavior.[74]

While critics are quick to argue that any of big data's benefits are underwhelming when weighed against potential harms, [75] the federal government can demonstrate the benefits of big data in a way that protects and promotes privacy. In particular, it can do this by supporting the development of big data tools and augmented accountability mechanisms across government. For example, the Federal Chief Information Officer (CIO) Council is well-positioned to ensure the federal government can maximize the potential of big data with an eye toward privacy protection.[76] The government can also do more to provide additional definition and framing as to how privacy principles like transparency and accountability can be used to alleviate concerns about big data.

Demonstrating how big data can be used to improve government function and services is only one component of the government's role, however. While the CIO Council can coordinate action within government, IPTF, again, can also bring agencies together to advance public policy that promotes big data.[77] Agencies can do more to support and highlight how big data is being used to benefit consumers in the private sector. Public/private partnerships can advance the necessary work to support innovation and advance privacy. IPTF should continue its work convening stakeholders and advancing further discussion about big data.

(3) What issues are raised by the use of big data across jurisdictions, such as the adequacy of current international laws, regulations, or norms?

As previously mentioned, FPF believes that big data may require a focus on accountability mechanisms to ensure responsible data stewardship across borders. To this end, FPF has advocated strengthening the existing US-EU Safe Harbor agreement and believes its continuation should be a

[72] Rebecca Sausner, *Warren's CFPB Embraces Big Data*, AM. BANKER (Dec. 1, 2010), http://www.americanbanker.com/btn/23_12/warrens-cfpb-embraces-big-data-1029410-1.html.
[73] Karuna Mintaka Kumar, *CFPB Tangles with Bankers over Big Data*, PYMNTS (July 19, 2013), http://www.pymnts.com/uncategorized/2013/cfpb-tangles-with-bankers-over-big-data/.
[74] *E.g.*, Op-Ed, *Consumer Financial Cover-Up*, WALL ST. J. (Mar. 17, 2014), http://online.wsj.com/news/articles/SB10001424052702303795904579431484040822904 ; Carter Dougherty, *Richard Cordray and the CFPB Are Monitoring Your Banking Habits*, BLOOMBERG BUSINESSWEEK (Apr. 25, 2013), http://www.businessweek.com/articles/2013-04-25/richard-cordray-and-the-cfpb-are-monitoring-your-banking-habits.
[75] For a discussion critiquing the benefits of big data, *see* Paul Ohm, *The Underwhelming Benefits of Big Data*, 161 U. PA. L. REV. ONLINE 339 (2013), http://www.pennlawreview.com/online/161-U-Pa-L-Rev-Online-339.pdf.
[76] CIO.gov, https://cio.gov/about/ (last visited Mar. 15, 2014).
[77] Nat'l Telecomm. & Info. Admin., Internet Pol'y Task Force, http://www.ntia.doc.gov/category/internet-policy-task-force (last visited Mar. 26, 2014).

top priority in the context of international data transfers.[78] More recently, the United States, in conjunction with representatives from the 21-nation Asia Pacific Economic Cooperation (APEC) and the EU's Article 29 Working Party announced the endorsement of a common referential.[79] This jointly-endorsed document identifies points of commonality under the APEC Cross Border Privacy Rules (CBPR) System and the EU's system of Binding Corporate Rules (BCRs). Each of these initiatives incorporates accountability mechanisms and clearly demonstrates the concept's utility in the context of international data transfers.

The Administration has previously committed to pursuing international interoperability through the mutual recognition of commercial data privacy frameworks that incorporate both effective enforcement and accountability mechanisms.[80] However, realizing the full potential of interoperability requires sustained senior-level engagement at the Department of Commerce. FPF urges the United States to use all levers of diplomatic, policy and regulatory activities across a range of international venues to achieve interoperability of these frameworks and consensus on their applicability in the era of big data.

Conclusion

Big data presents many benefits and potential risks. A thoughtful, balanced analysis of the value choices now at hand is essential. The Administration's efforts to convene thought leaders have produced many fruitful conversations, and more are needed. At the same time, it will be essential that the Administration provide transparency and a clear plan of action to all stakeholders moving forward.

Big data offers the United States a great opportunity to provide global leadership on promoting innovation – and protecting privacy. It also presents a challenge, but we have the privacy principles and frameworks needed to thoughtfully address that task.

FPF thanks the White House Office of Science and Technology Policy for considering these Comments, and we look forward to further engagement and collaboration on the issue of big data.

Sincerely,

Jules Polonetsky
Director and Co-Chair
Future of Privacy Forum

Christopher Wolf
Founder and Co-Chair
Future of Privacy Forum

Josh Harris
Policy Director
Future of Privacy Forum

Joseph Jerome
Policy Counsel
Future of Privacy Forum

[78] FUTURE OF PRIVACY FORUM, THE US-EU SAFE HARBOR: AN ANALYSIS OF THE FRAMEWORK'S EFFECTIVENESS IN PROTECTING PERSONAL PRIVACY (2013), http://www.futureofprivacy.org/wp-content/uploads/FPF-Safe-Harbor-Report.pdf

[79] Joint work between experts from the Article 29 Working Party and from APEC Economies, on a referential for requirements for Binding Corporate Rules submitted to national Data Protection Authorities in the EU and Cross Border Privacy Rules submitted to APEC CBPR Accountability Agents (Mar. 7, 2014), *available at* http://www.apec.org/~/media/Files/Groups/ECSG/20140307_Referential-BCR-CBPR-reqs.pdf.

[80] WHITE HOUSE BLUEPRINT, *supra* note 9, at 31.