



October 31, 2014

Federal Trade Commission  
Office of the Secretary  
Room H-113 (Annex X)  
600 Pennsylvania Avenue NW  
Washington, DC 20580

**Re: Comments for *Big Data: A Tool for Inclusion or Exclusion?***

Dear Sir/Madam:

The Future of Privacy Forum (FPF) is a think tank seeking to advance responsible data practices and includes leaders in business, academia, and consumer advocacy. At the Federal Trade Commission's (FTC) public Workshop exploring the use of big data and its impact on American consumers, Christopher Wolf, FPF Co-Chair and Founder, and Peter Swire, Senior Fellow, addressed some of the challenges and opportunities around big data. FPF now submits these additional comments on how to frame ethical conversations about big data moving forward. The path forward calls for the establishment of firmer frameworks around the use of data. As Chairwoman Ramirez suggested in her opening remarks, big data requires a serious conversation about "industry's ethical obligations as stewards of information detailing nearly every facet of consumers' lives."<sup>1</sup>

**I. Defining Big Data**

One key challenge moving forward is clarifying what we mean we are discussing "big data." As it stands, the term often means different things to different audiences in different contexts.<sup>2</sup> At the FTC's Workshop, for example, big data was separately understood as a "socio-technical phenomenon"<sup>3</sup> or as the remnants of our "digital exhaust."<sup>4</sup> It was discussed in relation to the activities of data brokers<sup>5</sup> as well as the broad use of social media.<sup>6</sup>

---

<sup>1</sup> FEDERAL TRADE COMMISSION, *BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION*, TRANSCRIPT at 13 (Sept. 15, 2014), [http://www.ftc.gov/system/files/documents/public\\_events/313371/bigdata-transcript-9\\_15\\_14.pdf](http://www.ftc.gov/system/files/documents/public_events/313371/bigdata-transcript-9_15_14.pdf) (statement of Edith Ramirez, FTC Chairwoman) [hereinafter *BIG DATA EVENT*].

<sup>2</sup> *What Is Big Data*, DATASCIENT@BERKELEY BLOG (Sept. 3, 2014), <http://datascience.berkeley.edu/what-is-big-data/>.

<sup>3</sup> *BIG DATA EVENT*, *supra* note 1, at 29.

<sup>4</sup> *Id.* at 117.

<sup>5</sup> *Id.* at 37.

<sup>6</sup> *Id.* at 113.

A traditional definition among data scientists is to discuss big data in terms of volume, velocity, and variety.<sup>7</sup> While this definition describes the phenomenon of big data, it does not directly explain the policy challenge posed by big data for privacy. Many of the contexts in which “big data” is used to describe a data use present no new difficulties or privacy concerns, or can be directly addressed through existing applications of the Fair Information Practice Principles (FIPPs). A broad definition of big data can obscure the reality that data uses cover a wide-range of different data practices, which should be evaluated by policymakers on a case-specific basis.

We propose cabining our discussion of big data to those practices that specifically challenge traditional applications of the FIPPs. The benefit of increasing amounts of data is the capacity for organizations and industry to harness existing information in novel ways to extract new insights and new forms of value.<sup>8</sup> When this is done to improve health care outcomes, provide personalized education, or ensure workplace diversity, these innovative uses of data can be curtailed by the rigid application of existing privacy protections.<sup>9</sup> It is at this point where data can be used for such great good that it stresses privacy constraints – or other existing legal frameworks – that big data becomes a policy challenge.

The FIPPs do have a degree of flexibility built into their application, and at different times, different principles have been emphasized ranging from the rights of individuals to the obligations of data collectors. Still, some of the most important FIPPs can be challenged by ubiquitous data. Some of key principles that can be strained by innovative data uses include: (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.<sup>10</sup> While the FIPPs should remain the foundation for our practices around privacy, their rigid application may not be appropriate in all instances going forward.

## II. Assessing the Benefits of Big Data

We have previously explained how these principles are stressed by big data,<sup>11</sup> and the challenge moving forward should be to better calibrate and augment the FIPPs in ways that both protect privacy and encourage innovative uses of data. In many cases, risk mitigation strategies can enable benefits while minimizing risks. In many cases, new uses may not create new risks. But when

---

<sup>7</sup> *Id.* at 15. Solon Barocas described big data by explaining that “the volume of data is exploding, that the velocity at which the data is accumulated is increasing, and the variety of formats of data is also likewise proliferating.”

<sup>8</sup> See VIKTOR MAYER-SCHONBERGER & KENNETH CUKIER, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK* (2013).

<sup>9</sup> Compare *BIG DATA: A TOOL FOR FIGHTING DISCRIMINATION AND EMPOWERING GROUPS*, FUTURE OF PRIVACY FORUM & ANTI-DEFAMATION LEAGUE (2014) with *BENEFIT-RISK ANALYSIS FOR BIG DATA*, FUTURE OF PRIVACY FORUM (2014), available at <http://www.futureofprivacy.org/2014/09/11/a-path-forward-for-big-data/>.

<sup>10</sup> 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>.

<sup>11</sup> See, e.g., Comments of the Future of Privacy Forum RE: Big Data RFI, at 3-7 (Mar. 2014), available at <http://www.futureofprivacy.org/wp-content/uploads/OSTP-Big-Data-Review-Comments.pdf>.

data innovation and privacy directly collide, FPF supports the development and expansion of rigorous data benefit assessments to help ensure ethical decisions are made..

At the Workshop, there was considerable discussion about the need to better understand the benefits and costs of data use.<sup>12</sup> While industry and privacy professionals have experience at assessing costs, the potential benefits of big data are less understood. Organizations must be guided by a risk-benefit analysis that takes into account exactly how the benefits derived from the use are distributed. FPF has proposed that entities develop procedures to assess the “raw value” of a data project that take into account (1) the nature of a project, (2) its potential beneficiaries, and (3) the degree to which that beneficiaries would actually benefit from the use of information.<sup>13</sup>

Benefit analysis may be an inherently subjective task, and one large challenge moving forward is determining who ultimately gets to decide when the benefits of big data trump any potential risks. Industry is increasingly turning to not just privacy officers and other privacy compliance professionals, but also outside advisory boards, internal committees, and other accountability processes to try to address the privacy challenges – and larger ethical questions – posed by big data.<sup>14</sup> These institutions can range from highly independent, publicly visible panels to internal organizations staffed by corporate officers. These sorts of review processes echo the development of Institutional Review Boards (IRBs) in response to ethical questions in the field of human subject testing, and provide an essential accountability mechanism for organizations that are struggling to balance the benefits and risks of data use.<sup>15</sup> Strong governance regimes, including ethical codes, staff, and adequate training, should be a prerequisite before any organization launches big data initiatives.<sup>16</sup>

### **III. Considerations Moving Forward**

In addition to establishing procedures and institutions to better consider the risks and benefits of big data, more work is needed to address where legal restrictions already cover potential big data uses and what ethical frameworks ought to apply in their absence. Identifying and understanding the harms presented by big data pose a particular challenge. Many of the new risks commonly associated with big data, such as filter bubbles or “creepy” inferences, are not easily mapped to any traditionally recognized harms.<sup>17</sup> Increasingly, big data calls for not just a better-accepted definition of the term itself but also clearer understanding of the contours of its harms. For example, there is a big difference between the potential harms from targeted online advertising derived from inferences derived from big data than there is from eligibility determinations based on data analytics, and different responses to the potential harms are needed. This is most certainly not an area for “one size fits all” solutions. Moreover, an overly inclusive classification of harms may lead to restrictions on the uses of big data that inhibits beneficial uses. This is

---

<sup>12</sup> See, e.g., BIG DATA EVENT, *supra* note 1, at 233, 237.

<sup>13</sup> See BENEFIT-RISK ANALYSIS FOR BIG DATA, *supra* note 9.

<sup>14</sup> Jules Polonetsky & Omer Tene, *Facebook Calls in the Philosophers*, THE HILL (Oct. 2, 2014), <http://thehill.com/blogs/pundits-blog/technology/219620-facebook-calls-in-the-philosophers>.

<sup>15</sup> See, e.g., Comments of the Future of Privacy Forum, Re: Big Data: A Tool for Inclusion or Exclusion? Workshop, Project No. P145406 (Aug. 2014), [http://www.ftc.gov/system/files/documents/public\\_comments/2014/08/00027-92420.pdf](http://www.ftc.gov/system/files/documents/public_comments/2014/08/00027-92420.pdf).

<sup>16</sup> KPMG, *Navigating Big Data's Privacy and Security Challenges 5* (2014), <http://www.kpmg-institutes.com/content/dam/kpmg/advisory-institute/pdf/2014/big-data-privacy-security.pdf>.

<sup>17</sup> See BENEFIT-RISK ANALYSIS FOR BIG DATA, *supra* note 9, at 3.

what Chris Wolf referred to at the Workshop as the potential “throwing the baby out with the bathwater” effect that should be avoided.<sup>18</sup>

While issues about potential unfair discrimination are driving much of the conversation about big data, efforts to provide additional clarity about what constitutes unfair discrimination through big data are needed. The White House Big Data Report specifically called for consultative processes to evaluate how big data interacts with existing privacy laws, and the FTC should play an active role as the NTIA continues its exploration of how data collection and use impact consumer privacy.<sup>19</sup> Industry also needs a better understanding of what activities are currently prohibited under existing federal laws. A variety of anti-discrimination laws, including the Equal Credit Opportunity Act and other fair lending laws, already provide a number of protections that are relevant in the context of big data when information is used for a variety of lending, housing, and employment purposes. The FTC should work together with other federal agencies to define the scope of these laws and provide additional guidance by identifying real-world examples that violate existing law.

In the meantime, businesses should take proactive measures both to protect privacy and to alleviate consumer’s concerns that data be used ethically and responsibly. While privacy professionals and other advisory boards can ensure ethical conversations are had about big data at a macro level, new consumer controls and education efforts can directly benefit individual consumers. For example, the Workshop highlighted the particular challenges that can arise from inaccurate information about a consumer. Ensuring that consumers are not miscategorized – and providing them the tools to remedy any error – will be essential to establishing trust in the use of big data. Existing fair lending laws suggest that organizations benefit from the implementation of rigorous compliance mechanisms, self-testing procedures, and other proactive measures.<sup>20</sup>

Better data may also be an important tool to ward against unfair discrimination. Fair lending enforcement has long been constrained by data limitations, but these data limitations are becoming easier to address due to increasing amounts of consumer data.<sup>21</sup> Further, the benefits of big data may not be distributed equitably if certain individuals and communities are inadvertently excluded from having their information be part of the “big data” dataset.<sup>22</sup> A “data divide” could exacerbate inequalities and, if certain individuals and groups are missing from a dataset, their problems could be overlooked despite benefits achieved elsewhere.<sup>23</sup>

While it will be important that everyone’s collective data is used to ensure big data serves as an inclusive tool, it is also essential that tools to protect individual privacy are also advanced. De-identification will be an important factor to protect privacy with the use of big data. The European Commission has identified the need for guidance on de-identification as an important element to promote privacy and security across a range of public-private partnerships on big

---

<sup>18</sup> BIG DATA EVENT, *supra* note 1, at 224, 250.

<sup>19</sup> WHITE HOUSE EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 61-65 (May 2014), [http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf).

<sup>20</sup> Peter Swire, Lessons from Fair Lending Law for Fair Marketing and Big Data (Sept. 2014), [http://www.futureofprivacy.org/wp-content/uploads/FairMarketingLessons\\_WhitePaperFTC.pdf](http://www.futureofprivacy.org/wp-content/uploads/FairMarketingLessons_WhitePaperFTC.pdf).

<sup>21</sup> *See Id.* at 10-11.

<sup>22</sup> Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55 (2013).

<sup>23</sup> Daniel Castro, Ctr. for Data Innovation, *The Rise of Data Poverty in America* (2014), <http://www2.datainnovation.org/2014-data-poverty.pdf>.

data.<sup>24</sup> Policy debates about the efficacy of de-identification must recognize that different protections and de-identification procedures that can exist across a spectrum of different data types in order to focus conversations on practical solutions.

Big data provides an important catalyst for conversations about privacy and ethics. The FTC is well positioned to lead those conversations, and to encourage industry to embrace new processes for addressing these challenges while ensuring the benefits of big data. FPF thanks the Commission for this opportunity to submit comments and looks forward to future engagement with the FTC on this topic.

Sincerely,

Jules Polonetsky  
Executive Director and Co-Chairman  
Future of Privacy Forum

Christopher Wolf  
Founder and Co-Chairman  
Future of Privacy Forum

---

<sup>24</sup> Press Release, Frequently asked questions: Public-Private Partnership (PPP) for Big Data, European Commission (Oct. 13, 2014), [http://europa.eu/rapid/press-release\\_MEMO-14-583\\_en.htm](http://europa.eu/rapid/press-release_MEMO-14-583_en.htm).