KEEPING THE INTERNET
OPEN • INNOVATIVE • FREE

www.cdt.org

CENTER FOR DEMOCRACY
& TECHNOLOGY

1634 Eye Street, NW
Suite 1100
Washington, DC 20006

March 31, 2014

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

Re: Big Data Study

The Center for Democracy & Technology (CDT) is pleased to submit these comments in response to the Office of Science and Technology's Request For Information (RFI) on the implications of big data.

Our comments focus on the continued value of the Fair Information Practice Principles (FIPPs) as the best available framework for addressing the privacy implications of big data practices; the possibility of technical measures, such as de-identification, to safeguard privacy; and the need for immediate reform of current laws, including the Electronic Communications Privacy Act (ECPA). We respond to each of the questions posed in the RFI in turn below.

> *(1) What are the public policy implications of the collection, storage, analysis, and use of big data? For example, do the current U.S. policy framework and privacy proposals for protecting consumer privacy and government use of data adequately address issues raised by big data analytics?*

In our view, big data involves the collection of vast amounts of data from a growing number and variety of sources, combined with powerful analytic techniques that promise to extract useful insights applicable to a range of business and social problems. While the U.S. has a concept of privacy, expressed in the Fair Information Practice Principles (FIPPs) and the Administration's Consumer Privacy Bill of Rights, that can be used to address and mitigate the privacy implications of big data, that concept has not been comprehensively implemented in U.S. law. To the contrary, even before the emergence of big data, it was widely recognized that U.S. law fails to provide adequate privacy protection in the face of the digital revolution.[1] The advent of big data should add urgency to the goal of updating U.S. laws, both for

[1] Testimony of Ari Schwartz, Center for Democracy & Technology, Senate Committee on Commerce, Science, and Transportation, Subcommittee on Interstate Commerce, Trade, and Tourism, "Reauthorization of the Federal Trade Commission" (Sept. 12, 2007), *available at* https://www.cdt.org/files/pdfs/20070912schwartz-testimony.pdf.

businesses and government, to establish comprehensive baseline privacy legislation and stronger standards controlling governmental access.[2]

Data is being collected about individuals in a growing number of ways – when they browse the Internet and use online services,[3] through their electrical energy smart meters, through mobile applications installed on smartphones,[4] through systematic monitoring of their Internet usage by their ISPs,[5] and by tracking their movements in a variety of ways,[6] among other methods. The Internet of Things will vastly magnify the potential for data collection.[7]

The use of this data to compile profiles and to make decisions about individuals raises fundamental issues of fairness. In theory, big data analytics could be used to classify individuals based on race, ethnicity, gender, national origin, age, sexual orientation, or other suspect classes.[8] Big data analytics could also be used in many ways to widen existing power disparities between companies and consumers, by more accurately determining that the precise price that any individual may be willing to pay for a specific commodity or service.

Even before analytic techniques are applied to make decisions about people, the *collection* of data implicates privacy interests.[9] By collecting vast sets of data, companies and governments open themselves up to risk of data breach, unintended exposure, and internal misuse. As entities amass larger databases of information that may be linked to individuals, those databases become tempting

---

[2] Our comments focus on big data applications that involve data about individuals or that draw inferences about individuals. We recognize that there are many big data applications (such as assessing the environmental conditions in a field of corn or the functioning of a jet airplane engine) that do not involve personally identifiable or re-identifiable data.

[3] Testimony of Justin Brookman, Center for Democracy & Technology, Senate Committee on Commerce, Science, and Transportation, "A Status Update on the Development of Voluntary Do-Not-Track Standards" (Apr. 24, 2013), *available at* https://www.cdt.org/files/pdfs/Brookman-DNT-Testimony.pdf.

[4] G.S. Hans, *Lookout's Open Source Privacy Policy Could Change the Game on Mobile App Transparency* (Mar. 27, 2014), *available at* https://www.cdt.org/blogs/gs-hans/2703lookouts-open-source-privacy-policy-could-change-game-mobile-app-transparency.

[5] G.S. Hans, *Should Your ISP Monitor What You Do With Your Internet Service?* (Aug. 13, 2013), *available at* https://www.cdt.org/blogs/gs-hans/1308should-your-isp-monitor-what-you-do-your-internet-service.

[6] Comments of Center for Democracy & Technology to Federal Trade Commission, February 2014 Workshop on Mobile Device Tracking (Mar. 19, 2014), *available at* https://www.cdt.org/files/pdfs/cdt-mobile-device-tracking-comments.pdf.

[7] Comments of Center for Democracy & Technology to Federal Trade Commission, November 2013 Workshop on "Internet of Things" (Jan. 10, 2014), *available at* https://www.cdt.org/files/pdfs/iot-comments-cdt-2014.pdf

[8] *See, e.g.,* Press Release, The Leadership Conference, "Civil Rights Principles for the Era of Big Data", http://www.civilrights.org/press/2014/civil-rights-principles-big-data.html.

[9] Justin Brookman & G.S. Hans, *Why Collection Matters: Surveillance as a De Facto Privacy Harm*, FUTURE OF PRIVACY F., *available at* http://www.futureofprivacy.org/wp-content/uploads/Brookman-Why-Collection-Matters.pdf

targets for malicious third parties seeking to gain unauthorized access. Depending on what information is contained within those databases, and how that information is protected (through de-identification, encryption, or other methods), the effects of a data breach could be catastrophic. As recent high-profile data breaches have demonstrated, sensitive personal and financial data can, in the event of a breach, become accessible to unauthorized third parties and can result in real-world consumer harm, such as identity theft.[10] Therefore, businesses and governments that collect data about individuals should limit their collection practices and only collect data necessary for specific uses.

The current American legal regime does not adequately protect consumer privacy. At present, there are a patchwork of laws, including the FTC Act, the Children's Online Privacy Protection Act, and the Video Privacy Protection Act, that provide varying degrees of privacy protection, but no comprehensive privacy legislation. CDT has long called for Congress to pass baseline privacy legislation,[11] and we have supported the proposals put forward by Congress,[12] the White House,[13] and the Federal Trade Commission.[14] The advent of big data should not be a distraction from this unfinished business; to the contrary, the increased surveillance, analytical and data retention technologies that make Big Data possible should spur the adoption of comprehensive baseline federal privacy legislation.

The FIPPs as a Framework to Protect Privacy in the Era of Big Data

CDT believes that the Fair Information Practice Principles (FIPPs) provide a robust framework to promote the protection of individual privacy interests in the era of big data. For the past thirty years, the dominant concept of information privacy has been expressed in the FIPPs. The Obama Administration adopted

---

[10] G.S. Hans, *Target and Neiman Marcus Testify on Data Breach – But What Reforms Will Result?* (Feb. 7, 2014), *available at* https://www.cdt.org/blogs/gs-hans/0702target-and-neiman-marcus-testify-data-breach—-what-reforms-will-result.

[11] Testimony of Justin Brookman, Center for Democracy & Technology, House of Representatives Committee on Energy and Commerce, Subcommittee on Commerce, Manufacturing, and Trade, "Balancing Privacy and Innovation: Does the President's Proposal Tip the Scales?" (Mar. 29, 2012), *available at* https://www.cdt.org/files/pdfs/Justin-Brookman-privacy-testimony.pdf

[12] Press Statement, Center for Democracy & Technology, CDT Statement on Release of Draft Consumer Privacy Bill: the Best Practices Act (Jul. 19, 2010), *available at* https://www.cdt.org/pr_statement/cdt-statement-release-draft-consumer-privacy-bill-best-practices-act.

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

[14] Statement of Edith Ramirez, Federal Trade Commission, Senate Committee on the Judiciary, "Privacy in the Digital Age: Preventing Data Breaches and Combating Cybercrime" (Feb. 4, 2014), *available at* http://www.ftc.gov/system/files/documents/public_statements/prepared-statement-federal-trade-commission-privacy-digital-age-preventing-data-breaches-combating/140204datasecuritycybercrime.pdf.

the FIPPs as the basis for its Consumer Privacy Bill of Rights in February 2012.[15] Many have argued that big data fundamentally challenges the FIPPs framework. In CDT's view, it is *not* inevitable that big data will overwhelm traditional concepts of privacy. Many of the issues now being cited in connection with big data are actually longstanding concerns (for example, the limitations of notice and consent). Many of the solutions being put forth by academics and others draw upon or echo elements of the traditional FIPPs framework.

For example, many of Paul Schwartz's recommendations sound very similar to core FIPPs.[16] Schwartz recommends, for example, that a company using big data analytics "should develop reasonable mitigation processes and reasonable remedies as appropriate when analytics lead to decisions that harm individuals," which sounds like the redress element of the individual participation FIPP and the accountability FIPP. Likewise, echoing the data quality FIPP, he recommends that a company "should engage in decision-making based on analytic output that is reasonably accurate." At another point, Schwartz recommends that, based on ongoing review and revision of their analytics practices, companies "should only use information that is predictive," which restates the data quality principle to emphasize the reliability of outcomes.

While initially dismissing the traditional privacy framework, Christopher Kuner and co-authors also end up endorsing the FIPPs that focus on outcomes.[17] Given big data's role in decision-making about individuals, they state, "issues such as the accessibility, accuracy and reliability of data may matter as much or maybe more than privacy" (by "privacy," the authors seem to mean collection and use limitations). Of course, accessibility, accuracy and reliability have always been key FIPPs. Also, it is noteworthy that Kuner et al., after severely criticizing the consent model, conclude that there remains a proper role for individual consent, further illustrating how compelling the FIPPs framework is.

---

[15] *See* WHITE HOUSE, *supra* note 13, at 1. The FIPPs were first articulated in both the U.S. and in Europe in the early 1970's and quite rapidly became the focus of privacy policy development on both sides of the Atlantic and, after their adoption by the OECD, globally. *See* Robert Gellman, *Fair Information Practices: A Basic History*, http://bobgellman.com/rg-docs/rg-FIPShistory.pdf. The Department of Homeland Security adopted a version of the FIPPs as the foundation for privacy policy and implementation at DHS in 2008. Hugo Teufel III, Department of Homeland Security, *Privacy Policy Guidance Memorandum* (Dec. 29, 2008) *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf. Here, we use both the language of the Administration's FIPPs, which focused on the consumer context, and the DHS FIPPs, which focused on government practices. The congruence between the Administration's formulation and DHS's shows the consistency in the understanding of information privacy in the U.S.

[16] See Paul Schwartz, *Data Protection Law and the Ethical Use of Analytics*, Privacy & Security Law (Jan. 10, 2011), and Paul Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2055, 2096 (2004).

[17] Christopher Kuner, Fred H. Cate, Christopher Millard, & Dan Jerker B. Svantesson, *The Challenge of "Big Data" for Data Protection*, 2 INT'L DATA PRIVACY L. 49 (2012).

Omer Tene and Jules Polonetsky note that the FIPPs have a certain adaptability that allows adjustments in emphasis among the various principles.[18] They base their solution on "re-craft[ing] transparency obligations and access rights to make them more useful in practice." They note that "[t]raditional transparency and individual access mechanisms have proven ineffective." However, rather than proposing to replace transparency and access, they call for more effective implementation of these FIPPs, which they argue will both better protect individuals and unleash the power of big data:

> "If organizations provide individuals with access to their data in usable format, creative powers will be unleashed to provide users with applications and features building on their data for new innovative uses. In addition, transparency with respect to the logic underlying organizations' data processing will deter unethical, sensitive data use and allay concerns about inaccurate inferences."

It appears that the FIPPs framework is more durable than many have recently assumed. Even in calling for new approaches to privacy in response to big data, key experts have reverted to concepts that are part of the FIPPs. There is a certain power in this correlation between traditional data protection concepts and new ideas about privacy and big data. Among other things, the correlation offers a response to the paralysis of privacy policy that big data seemed to portend. It is not necessary to have a full rethink of privacy.

Indeed, in our view, rather than tolling the death knell of privacy, the big data phenomenon could be leveraged to spur development of the workable and effective privacy framework that has long been lacking. We believe that any comprehensive federal privacy legislation should use the FIPPs as an organizing framework, and that, in the interim, companies should use the FIPPs as a self-regulatory tool to protect their customer's privacy interests. Below, we discuss the FIPPs in turn and highlight their relevance to big data.

Purpose Specification and Use Limitation (Respect for Context)

It is often said that big data techniques involve the use of data in unanticipated ways. Nevertheless, purpose specification and use limitation are two closely related principles that remain vital to protecting individual privacy. With respect to consumers, the Administration's Privacy Bill of Rights well describes the two principles when it says, "Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data." Even in the era of big data, purpose specification should be a crucial first step in any system design, requiring entities to detail on what grounds they will collect data and the uses that they plan for it. The use limitation principle requires entities to follow through on the delineated uses and refrain from using the collected data for undisclosed purposes.

[18] Omer Tene and Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239 (2013).

Limitations on the collection of data are vitally important in a world in which it is becoming less and less expensive to collect increasing amounts of data from a variety of devices. Individual privacy interests are implicated at the point of collection, because of the variety of risks that databases are subject to. When any entity collects data about individuals, that data can be subject to internal misuse, changes in company practices, or data breaches.[19] Some have argued that relying on use limitations would be sufficient to protect privacy, but the threats to privacy arise long before an entity actually uses the data. Use limitations, while important, cannot protect against all possible threat models. As a result, purpose specification, which provides both a basis for and limits on the collection of information, is a vital element to protecting individual privacy interests. It is directly linked to other principles, including minimization (focused collection) and transparency. Companies engaged in big data analytics should be sure to detail the purposes for which they collect information in order to demonstrate their commitment to protecting consumers and their privacy interests. Use limitations are also important. Companies must confine their uses of data to the purposes disclosed to consumers. If the company plans to share data collected with a third party, that sharing should be disclosed to consumers in advance, as should the third party's uses (e.g. analytics).

Especially in the big data context, entities collecting personal information could very well develop new uses of data in future years that are loosely (if at all) related to the uses that the data was originally collected for. If that happens, entities must at the very least provide transparency about those new uses before they begin. Entities holding data should consider whether the new uses can be performed with de-identified data. They should carefully weigh the potential adverse consequences that may befall individuals from the use of such data and design their programs to avoid such consequences or ensure that they are reliable and justified. However, if data custodians conclude that the new uses can only be performed with identifiable data, and are not contextually related to the purposes for which the data was originally collected, they must seek new consent for those new uses. User expectations – and the potential for user *surprise* – are important indicia for whether a new purpose is contextually related to an older one.

Transparency

It is almost certain that the public has very low awareness of the uses currently made of data – much less potential future uses. Therefore, the onus is on the *companies, governmental entities, and others collecting and using data about individuals* to disclose what uses they are making and plan to make, and, when they come up with new uses, to disclose them. Both the private sector and the government will have to educate the public on what big data actually means and why entities are employing it. By being transparent about their collection, use, and retention practices of data, companies, government agencies and other entities will both create better public awareness of their practices and increase public trust.

[19] *See* Brookman & Hans, *supra* note 9.

The limitations of notice have long been recognized. Whether it is corporate privacy policies or Privacy Act System of Records Notices, individuals are unable to sift through the massive volume of verbiage to determine which is relevant. But at the very least, companies and government agencies must make information about all their practices available to the public in some form – whether in a privacy policy, in terms of service, in the statutes and guidelines defining governmental collection authorities, or in other forms of detailed disclosure. The ability for the public to access information on corporate and government practices is vitally important, both for educational purposes and to hold companies and government officials accountable when their public statements fail to correspond with their actual practices. The FTC should continue to undertake investigations and bring enforcement actions against companies that have not sufficiently described their data privacy practices.[20]

Individual Participation (Individual Control)

Related to the transparency principle, the individual participation principle urges companies to give individuals control over what personal data is collected from them and how it is used. The most obvious way that companies can do this is by allowing users to make decisions regarding what data gets collected, and what uses a company can make with that data. Especially where consumers purchased the devices that enable big data analytics, they should be in control over what data those devices collect and transmit to companies. Therefore, companies should solicit the participation of consumers when seeking to access the data that devices can provide.[21] Some data collection and retention can reasonably be done only on an opt-out basis – that is, unless the consumer affirmatively objects. Some – such as data collected and used only for reasonable and focused security purposes – should not be subject to individual control at all. However, certain sensitive categories of data should only be collected and retained with a consumer's informed permission. Information about medical conditions – or information about what users do inside their own homes – are examples of intensely personal information that should only be done on an opt-in basis.

The development of effective notice and consent regimes will play a vital role in enabling responsible big data regimes, as pervasive collection may allow businesses to create highly granular and comprehensive records for individual customers. Because more and more companies have the capacity to monitor users, these controls will in many cases need to be universal. For example, the Do Not Track mechanism has been proposed as an easy and effective way for consumers to express their choice to stop cross-site tracking in the online

[20] G.S. Hans, *Goldenshores Case Demonstrates Flaws in Current Mobile Privacy Practices* (Dec. 23, 2013), *available at* https://www.cdt.org/blogs/gs-hans/2312goldenshores-case-demonstrates-flaws-current-mobile-privacy-practices.

[21] A recent case involving LG TVs that broadcast viewer usage practices to the manufacturer highlights the need for empowering users to make the final say over how their devices behave. *See* Justin Brookman, *Eroding Trust: How New Smart TV Lacks Privacy by Design and Transparency* (Dec. 3, 2013), *available at* https://www.cdt.org/commentary/eroding-trust-how-new-smart-tv-lacks-privacy-design-and-transparency.

context. The online advertising industry should be encouraged to honor users' Do Not Track requests, and other industries should explore similar universal choice mechanisms to allow consumers to more effectively regulate the dissemination of their personal information.[22]

Effective consumer notification will be necessary. Customers may not even be aware what a business can collect from a computer, a mobile device, or wearable devices. Without adequate notice and consent provisions, customers who don't approve of what a particular business does won't be able to "vote with their feet" and choose another business with different practices. Companies should begin developing effective consent models now, rather than deploying them after they finalize their big data collection practices.

Security

The recent spate of high-profile data breaches emphasizes the need for strong security programs for all entities that collect data about individuals.[23] As the big data trend results in the increasing collection of data, businesses and governmental entities must create strong security programs – and monitor and update those programs – in order to protect data. Companies should be held accountable for failing to safeguard the data they maintain and should notify consumers of breaches as they occur in full compliance with current law. Although the FTC's ability to seek enforcement actions against companies for poor data security practices is currently being litigated, CDT thinks that the FTC currently has authority under Section 5 of the FTC Act to regulate data security, and we encourage the FTC to continue to bring enforcement actions against companies that have substandard data security programs.[24]

Data Minimization (Focused Collection)

Data minimization is closely related to data security. Collecting data without a clear (and disclosed) purpose in mind, or the failure to purge old data in accordance with reasonable minimization procedures, should be factors in evaluating whether an entity's data security practices were reasonable. As part of their security programs, companies, government agencies and other entities should implement specific retention periods for data, rather than retaining that information indefinitely. If entities implement minimization procedures and delete unnecessary, outdated, or irrelevant entries, fewer records will be accessible to unauthorized parties if and when a data breach occurs. By removing identifying

[22] For example, FTC Commissioner Julie Brill has launched an initiative, "Reclaim Your Name", that would educate users and empower them to assert control over their personal data. *See* Julie Brill, Op-Ed, *Demanding Transparency from Data Brokers*, WASH. POST (Aug. 15, 2013), *available at* http://www.washingtonpost.com/opinions/demanding-transparency-from-data-brokers/2013/08/15/00609680-0382-11e3-9259-e2aafe5a5f84_story.html.

[23] See Hans, *supra* note 10.

[24] G.S. Hans, *Data Security and Your Next Hotel Stay: How the FTC Encourages Strong Security Practices* (May 21, 2013), *available at* https://www.cdt.org/blogs/gs-hans/2105data-security-and-your-next-hotel-stay-how-ftc-encourages-strong-security-practice.

information and deleting data after it is no longer needed, companies will both protect their customers' security and promote consumer trust.

If a company retains data or shares it with a third party, it should consider anonymizing or pseudonymizing the data it provides in order to protect individual privacy. In its 2012 report on consumer privacy, the FTC set out the following standard to ensure that data is properly anonymized so that it cannot be "reasonably linked" to a particular consumer, computer, or device: "data is not 'reasonably linkable' to the extent that a company: (1) takes reasonable measures to ensure that the data is de-identified; (2) publicly commits not to try to re-identify the data; and (3) contractually prohibits downstream recipients from trying to re-identify the data." [25] CDT believes that this is an appropriate and viable standard for companies to implement to de-identify consumer data. By removing identifying information before sharing data, companies can take an affirmative step to protecting consumers even after the data is out of their direct control by reducing the likelihood that someone else can use the data for undisclosed purposes

Data Quality (Access and Accuracy)

Entities collecting and using data about individuals should also ensure that the data they use and retain is accurate, relevant, and complete. Because of the sensitive nature of data collected for big data purposes, it is vitally important for entities to ensure that their records are accurate. If a promotional offer was delivered to the wrong consumer or if records were not kept suitably secure, customers could become disturbed, inconvenienced, or vulnerable to inappropriate uses.[26]

Accountability and Auditing

In order to ensure that data collection and use practices are followed and security programs are properly implemented, entities must create internal oversight mechanisms and must be subject to external accountability. This will ensure that the practices that are nominally adopted are effectively followed, and will encourage public trust.

(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?

[25] FEDERAL TRADE COMM'N, *Protecting Consumer Privacy in an Era of Rapid Change* (February 2012), *available at* http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

[26] Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAGAZINE at MM 30 (Feb. 19, 2012), *available at* http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

As discussed above, uses of big data that classify individuals based on suspect classes and treat those individuals differently from the general public would raise the most public policy concerns. Because big data analytics are conducted without public knowledge or disclosure, it is difficult to identify when such classifications are being made. Therefore, the openness principle described above will be particularly important to determine when businesses are making such classifications so that consumers can make a more informed choice about what data they provide to businesses.

Government uses of big data also raise public policy concerns. While the Privacy Act regulates the government's ability to create and use databases, its provisions include multiple exceptions for agencies that engage in law enforcement and foreign intelligence. As a result, individuals whose records are collected and analyzed by the government may not be aware that such analysis is taking place. Increased transparency concerning government use of personal data in big data processes will help limit this type of use.

The use of algorithms to make determinations regarding how individuals should be classified, targeted, or marketed to may raise specific policy concerns. If companies make assumptions about what an individual wants and targets those individuals, those assumptions may be reinforced rather than challenged as inaccurate or outdated. For example, if a store targets individuals with advertisements or coupons for foods with high sugar, fat, or salt content, the store may undermine the customer efforts to adhere to a diet plan. Ryan Calo has described this issue as "digital market manipulation," arguing that such practices deserve attention from regulators.[27]

> *(3) What technological trends or key technologies will affect the collection, storage, analysis and use of big data? Are there particularly promising technologies or new practices for safeguarding privacy while enabling effective uses of big data?*

The trend towards increased collection of data from mobile devices, networked appliances, cars, wearable devices, and online services allows for more extensive big data analytics. Advances in computing power and storage capacity allow companies and government to retain more data for longer periods at reduced costs, and analyze massive datasets to an unprecedented degree.

The possibility of collecting all data on a persistent basis has given rise to dystopic fears of a world in which someone is always watching you, even in private spaces. The United States Constitution specifically identifies realms in which individuals have heightened privacy interests, and big data collection capabilities imperil those spaces. There have been many stories, from FTC enforcement actions to inappropriately targeted advertisements, that highlight how individual privacy has been compromised by systems that were designed to

[27] Ryan Calo, *Digital Market Manipulation*, forthcoming GEO. WASH. L. REV. (2014), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2309703.

constantly collect data without foresight. From laptop cameras spying on individuals during intimate moments,[28] to advertisements identifying a teenager as pregnant before she told her parents,[29] consumers have been threatened by invasive practices that could have been avoided had companies used more forethought in designing systems and releasing products.

Safeguarding privacy, therefore, is of increased importance given that there are more ways than ever before for privacy interests to be implicated by big data practices. Allowing individuals to have as much control as possible over their devices will be paramount in protecting those interests. When a consumer purchases a device that has the capacity to collect data, that consumer should have the ability to the extent possible to control that collection. Some practices that allow for big data analytics – such as mobile device tracking – rely upon data collection that individuals may not be aware is even happening, and may have few ways to prevent. For example, mobile device tracking typically uses a device's broadcast of a Media Access Control (MAC) address to track that device over time and make determinations based on the behavior of the device (and by implication, the device's owner). However, due to the system architecture of most devices, many individuals are not aware that their devices are broadcasting MAC addresses, and are not easily able to prevent it from happening without disabling Wifi and Bluetooth functionality.[30] When consumers purchase devices, they should be the ultimate arbiter of when data collection occurs, how it occurs, and with what frequency. Empowering users to have more control over their devices and the types of data that each device collects will allow individuals to proactively protect their privacy.

Companies should also limit the amount of data they collect. There has been a worrisome trend to focus on use limitations, rather than limitations on collection. However, for the reasons discussed above, there are multiple ways in which the collection of data implicates individual privacy interests. Companies should therefore make deliberate decisions on what types of data are collected and under which circumstances, rather than enabling all possible types of collection and relying upon use limitations to manage their databases. Use limitations are important, but must be coupled with collection limitations in order to create a FIPPs-compliant data management regime.

Some authors have proposed specific technical solutions, such as an increasing reliance on differential privacy as proposed by Cynthia Dwork.[31] Companies have also architected some big data analytic systems to be privacy protective; for example, Facebook created a double-blind system when partnering with

---

[28] G.S. Hans, *Laptop Spying Case Indicates More Aggressive FTC Stance on Privacy* (Oct. 9, 2012), *available at* https://www.cdt.org/blogs/gs-hans/0910laptop-spying-case-indicates-more-aggressive-ftc-stance-privacy.

[29] *See* Duhigg, *supra* note 26.

[30] *See* Comments of Center for Democracy & Technology, *supra* note 6.

[31] Cynthia Dwork, *Differential Privacy: A Survey of Results* (2010), *available at* http://www.cs.ucdavis.edu/~franklin/ecs289/2010/dwork_2008.pdf.

Datalogix in order to ensure that neither company could create a detailed profile based on online and offline data.[32] We at CDT hope that other companies will take steps to develop such technical solutions.

> *(4) How should the policy frameworks or regulations for handling big data differ between the government and the private sector? Please be specific as to the type of entity and type of use (e.g., law enforcement, government services, commercial, academic research, etc.).*

One major difference between policy frameworks for the government and the private sector is that the government is subject to the Constitution. Unfortunately, the case law and statutes have not kept pace with technological developments, and personal data held by third parties is inadequately protected against government access. A major challenge that needs to be addressed, and one where we urge the Administration to take a stronger position, is to ensure that the principles of the Fourth Amendment are extended to cover government access to digital data held by third parties.

Immediate reform is needed with respect to the content that individuals are storing with third party companies more than ever before. This includes emails, photos, address books and documents stored in the cloud.

Under an outdated law, this digital content is not adequately protected from government access. The Electronic Communications Privacy Act (ECPA) says that government agencies do not need a warrant— authorized by a judge and based on probable cause—to demand that third party service providers turn over the contents of their customers' emails and documents. A federal appeals court, in a decision we endorse, has held that ECPA is unconstitutional in this regard. Bi-partisan legislation is pending in both Houses of Congress to address this problem. The Administration should support enactment of S. 607 and H.R. 1852, with no carve-outs or exceptions for civil agencies.

The problem of government access also extends to metadata about communications. Cell site location data is one particularly revealing type of data and is automatically generated by mobile phones used by 91% of the U.S. population.[33] The government argues that it does not need a warrant to access consumers' mobile location data held by communications service providers. Of all the kinds of transactional data, location tracking information is one that clearly should be subject to the warrant protection. Again, bipartisan legislation is pending in both Houses of Congress to require government agencies to obtain a warrant before compelling service providers to disclose location tracking information, and the Administration should support that legislation.

---

[32] Jennifer Martinez, *Facebook's New Ad Partnership Stokes Privacy Concerns*, THE HILL (Sept. 26, 2012), *available at* http://thehill.com/blogs/hillicon-valley/technology/251287-facebooks-new-ad-tracking-partnership-stokes-privacy-concerns-.

[33] Lee Rainie, *Cell Phone Ownership Hits 91% of Adults*, FactTank: Pew Research Center (June 6, 2013), *available at* http://www.pewresearch.org/fact-tank/2013/06/06/cell-phone-ownership-hits-91-of-adults/.

At the root of concerns about government access in the era of big data is the so-called third party doctrine. If the Administration wants to do anything about big data and privacy it at least needs to acknowledge that the scope of the third party doctrine needs to be curtailed. Adopted long before digital technology had become essential to daily life and long before the outlines of the big data phenomenon were apparent, the third party doctrine says that individuals lose all constitutional privacy interest in data voluntarily disclosed to a third party. This doctrine is the basis of arguments that there is no constitutional privacy interest in documents stored in the cloud, in cell phone tracking information, or in records collected by the private sector about our daily activities, ranging from health to finances to travel to entertainment choices. It is the basis of the NSA telephony metadata program, the revelation of which helped prompt this review.

The third party doctrine is especially ill-suited to the era of big data, for it says that all of the big data collected by commercial entities about individuals is unprotected by the Constitution. Until the third party doctrine is addressed, government access issues will be left to a patchwork of statutes, many of which currently allow government access to highly sensitive data under a very weak standard.

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

Because of disparate international laws and regulations, multinational companies – whether they use big data analytics or not – need to comply with often contradictory regulations. The lack of a comprehensive U.S. privacy framework has made this particularly difficult. The European Union is debating a Data Protection Regulation (DPR) that may impose even further limits on the abilities of companies to conduct big data analytics.[34]

The lack of a baseline consumer privacy law in the U.S. makes it harder for U.S. companies and officials to argue credibly against overbroad or unworkable privacy regulations. The EU DPR was proposed in part to force American companies to institute stronger privacy protections, specifically in response to the fact that current American law does not place baseline requirements upon companies.[35] In our view, continuing U.S. inaction on consumer privacy contributes to proposals in Europe that would be unduly restrictive of the open Internet. Proposed European data localization requirements, for example, would not only negatively affect American businesses by increasing operating costs for companies or leading European users to migrate from American services to

[34] Justin Brookman, *Progress Made Revising EU Data Privacy Laws* (Nov. 12, 2013), *available at* https://www.cdt.org/blogs/justin-brookman/1211progress-made-revising-eu-data-privacy-laws.

[35] Kevin J. O'Brien, *Firms Brace for New European Data Privacy Law*, N.Y. TIMES (May 13, 2013), *available at* http://www.nytimes.com/2013/05/14/technology/firms-brace-for-new-european-data-privacy-law.html.

European analogues.[36] Data localization would also contribute to fragmentation of the open Internet. However, so long as the U.S. lacks a privacy law, countries may be attracted by such extreme measures, and countries outside Europe will likely defer to the European framework for consumer privacy protection when developing their own regulations. If Congress did pass baseline privacy legislation, it would signal to the world that data can be protected without the need for localization.

Even in the absence of baseline federal privacy legislation, businesses should adhere to practices that allow for user control, not only to promote individual privacy but also to increase the likelihood of complying with disparate international law. Technological changes and advancements should not fundamentally change human rights protections or the need for private locations in which individuals can exist unobserved. The rise in networked devices that can silently collect data in the home and mobile devices that can track the owner's location may imperil those private spaces, and companies should consider privacy – which has been considered an international human right – when designing their products and systems.

Conclusion

We thank OSTP for soliciting comments and for its workshop series on big data and its technical, social, and regulatory implications. Faced with the privacy and security risks inherent in big data practices, we believe the FIPPs are as relevant as ever and provide an exemplary framework for promoting individual privacy protections by both business and government. A FIPPs-based framework could address the key challenges of big data:

- provide protections for privacy while still enabling analytics to solve pressing business and social challenges;
- apply consistently across sectors yet still be flexible enough to respond to the particular risks to privacy posed by different applications;
- include mechanisms to hold accountable entities collecting and analyzing data; and
- provide incentives for the adoption of privacy-enhancing technical architectures/models for collecting and sharing data.

[36] Alison Smale, *Merkel Backs Plan to Keep European Data in Europe*, N.Y. TIMES (Feb. 16, 2014), at A6, *available at* http://www.nytimes.com/2014/02/17/world/europe/merkel-backs-plan-to-keep-european-data-in-europe.html.

Sincerely,

/s/
Nuala O'Connor
President & CEO

/s/
Jim Dempsey
Vice President for Public Policy

/s/
Justin Brookman
Director, Consumer Privacy Project

/s/
Joseph Lorenzo Hall
Chief Technologist

/s/
G.S. Hans
Ron Plesser Fellow

/s/
Runa A. Sandvik
Staff Technologist