CENTER FOR URBAN
SCIENCE+PROGRESS

1 MetroTech Center, 19FL
Brooklyn, New York 11201
tel: 646.997.0500
fax: 646.997.0560
web: cusp.nyu.edu

John Morris
National Telecommunications and Information Administration
U.S. Department of Commerce
1401 Constitution Avenue NW, Room 4725
Washington, DC 20230

August 5, 2014

RE: Big Data and Consumer Privacy in the Internet Economy, 79 Fed. Reg. 32714 (June 6, 2014)

Established April 23, 2012 as part of the City of New York's Applied Sciences NYC initiative, the Center for Urban Science + Progress (CUSP) at New York University is about the intersection of two simple but compelling themes: cities and data. Cities are where the people are; about half of humanity lives in urban environments today and that number will grow to 80 percent by the middle of this century. The ability to collect, transmit, store and analyze data is rapidly growing, and if properly acquired, integrated, and analyzed, big data can take governments beyond today's imperfect and often anecdotal understanding of cities to better operations, better planning and better policy.

While the focus in recent years has been on the exploitation of big data for commercial and national security purposes, CUSP believes data science has much to contribute to the public good. Data science can help address major questions about urban infrastructure, the urban environment, and the interactions of people with each other, with institutions, their interactions as organizations, as well as their interactions with the built and natural environments. Yet it is the ever finer temporal and spatial granularity of data about individuals and the increasing power of informatics tools to combine and mine these streams of data that stoke concerns about privacy and data access, particularly when these tools are in the hands of individuals or organizations whose interests are not perceived as being aligned with those of the data subjects. CUSP is dedicated to developing the big data tools that will help cities and the City of New York, in particular, become more productive, livable, equitable and resilient.

CUSP commends the Department of Commerce for providing a forum where big data developments and their impact on the Consumer Privacy Bill of Rights can be explored. Since CUSP submitted comments in response to the Office of Science & Technology Policy's "Big Data RFI" (79 Fed. Reg. 12251), CUSP is providing additional input on only 5 of the 20 questions included in this Request for Public Comment.

2. Should any of the specific elements of the Consumer Privacy Bill of Rights be clarified or modified to accommodate the benefits of big data? Should any of those elements be clarified or modified to address the risks posed by big data?

In their chapter *Big Data's End Run Around Anonymity and Consent* in the recently published book, *Privacy, Big Data and The Public Good: Frameworks for Engagement*, Solon Barocas and Helen Nissenbaum conclude that the transition to big data has revealed intractable challenges to both consent

and anonymity.[1] The logic underlying their position suggests that the individual control, transparency, respect for context, and focused collection elements of the Consumer Privacy Bill of Rights, alone, are not sufficient in the face of challenges posed by big data. These elements need to be considered within a responsible use framework.

Barocas and Nissenbaum argue that consent, as currently constructed, has little traction against the challenges of big data.[2] Instead, they propose that notice should be reserved for notable departures from the data subject's reasonable expectation and that consent should not be required for acceptable, expected uses. The burden should be on notice to describe clearly the violations of norms, standards, and expectations for which consent is being sought.[3]

When consent is sought, data subjects must be well informed of departures from expected behaviors and the scope of the consent must be consistent with the reasons data subjects have for acquiescing.[4] Examples include uses for which individuals are presumed to waive specific rights or for actions that promise significant benefits to others or to society at large.

This approach is not so different from that articulated by the Article 29 Data Protection Working Party in its opinion on legitimate interests.[5] In the European Union, the processing of personal data based on the controller's "legitimate interests" is one of the options available when consent is unobtainable or impracticable, such as when companies process big data. Using legitimate interests as the ground for data processing in the European Union requires a balancing on the one hand of the legitimate interests pursued by the data controller and on the other hand the interests and fundamental rights of the data subject; if the balancing falls in favor of the data subject, companies are not allowed to use the controller's legitimate interests as a legal ground for the processing of personal data. In a not so formalistic way, this is what Barocas and Nissenbaum are saying as well. A balancing approach must be employed. Actions that otherwise violate important norms, standards or expectations must be justified the organizations in possession of the data or by those carrying out the analyses.[6]

Big data analytics, in the simplest framing, comprise a two-phase process that includes knowledge discovery and application.[7] The knowledge discovery phase involves gathering data to be analyzed, getting it into a format that can be used, consolidating it for analysis, analyzing it to discover what it may reveal, and interpreting it to understand the processes by which the data was analyzed and how conclusions were reached. The product of the knowledge discovery phase is frequently an algorithm. Through application of derived algorithms, organizations make determinations upon which they can act, thus reaping the benefits of knowledge discovery.[8] In general, these two commonly recognized phases of big data analytics should be regarded as acceptable, expected behavior, and it should be assumed, in developing a reasonable workable framework, that data will be used for big data analytics. Indeed, the

---

[1] Solon Barocas and Helen Nissenbaum, *Big Data's End Run Around Anonymity and Consent* published in *Privacy, Big Data, and the Public Good: Frameworks for Engagement* at 45 (Cambridge Press 2014) [hereinafter *Barocas and Nissenbaum*].
[2] *Id.* at 58.
[3] *Id.* at 65.
[4] *Id.* at 65 – 66.
[5] Opinion 06/201 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/6/EC, available at nttps://www.huntonprivacyblog.com/files/201.0/wp217_en.pdf
[6] See *Barocas and Nissenbaum* at 65.
[7] *Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance*, Centre for Information Policy Leadership (February 2013) available at http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf.
[8] *Id.*

sharing by commercial organizations of data with researchers at colleges and universities in order that research for the public good can be undertaken, subject to well established Internal Review Board protocols,[9] should be considered consistent with norms, standards, and expectations. Consumer education about big data analytics should be offered, and all stakeholders, including industry, trade associations, consumer and privacy groups, researchers and government entities, should engage in these education efforts. In cases where consumers do not want their data used for big data analytics, they should be offered clear opportunities to opt out.[10]

When big data analytics are conducted in ways that deviate from important norms, standards, or expectations, organizations must be required to justify their actions as part of demonstrating that they are accountable organizations. Accountability, one of the principles of the Consumer Privacy Bill of Rights, requires companies to have measures in place to demonstrate to enforcement authorities and consumers upon request the development and implementation of appropriate assessments, policies and procedures consistent with the other principles in the Consumer Privacy Bill of Rights.

3. Should a responsible use framework, as articulated in Chapter 5 of the Big Data Report, be used to address some of the challenges posed by big data? If so, how might that framework be embraced within the Consumer Privacy Bill of Rights? Should it be? In what contexts would such a framework be most effective? Are there limits to the efficacy or appropriateness of a responsible use framework in some contexts? What added protections do usage limitations or rules against misuse provide to users?

CUSP agrees with the Big Data Report and the PCAST Report that the Consumer Privacy Bill of Rights should address the unique characteristics of big data and be revised to put a greater emphasis on a responsible use framework. The PCAST Report divides the seven principles embodied in the Consumer Privacy Bill of Rights into two categories: data holder obligations and consumer empowerments.[11] CUSP supports revisions to five of the seven Consumer Privacy Bill of Rights principles as suggested in the PCAST Report:[12]

Data Holder Obligations

- ***Respect for Context***: Data about an individual – however acquired – should not be used so as to cause adverse consequences or harms to that individual.
- ***Focused Collection***: This right should be about use rather than collection, and best practices should be emphasized to prevent inappropriate use of data during the data's whole life cycle, rather than depending upon de-identification.

Consumer Empowerments

- ***Individual Control***: Each company should take responsibility for conforming its uses of personal data to a personal profile designated by the consumer and made available to that company.
- ***Transparency***: The burden of conforming its uses of personal data to a consumer's stated personal-privacy profile should fall on the company.
- ***Access and Accuracy***: When the personal character of the data is apparent to a company by virtue of its use of the data, [the company's] obligation to provide means for correction of errors should be triggered.

---

[9] See CUSP Response to "Big Data RFI" at 6 – 7.

[10] Today they can opt out only through the Network Advertising Initiative Opt Out available at http://www.networkadvertising.org/choices/. In the future, CUSP recommends they can opt out as part of their personal privacy profile, see response to Question 4, pages 4 – 6, *infra*.

[11] PCAST Report at 43.

[12] PCAST Report at 44-45.

With respect to big data, the Consumer Privacy Bill of Rights appears more focused on collection and "notice and choice" rather than on use. The greatest potential for harm arises from the use of data, and given the volume, velocity, and variety of data being used, a responsible use framework should be more explicitly embraced within the Consumer Privacy Bill of Rights.

4. What mechanisms should be used to address the practical limits to the "notice and consent" model noted in the Big Data Report? How can the Consumer Privacy Bill of Rights' "individual control" and "respect for context" principles be applied to big data? Should they be? How is the notice and consent model impacted by recent advances concerning "just in time" notices?

As discussed above, "notice and consent" are considered to be consumer empowering. By limiting the applicability of "notice and consent" in the big data context, consumer empowerment is not being reduced. The way for consumers to be empowered is not to continue to burden them with the management of privacy for each company with which the consumer interacts by the "notice and consent" model. Rather the responsibility should be on each company to conform its uses of personal data to the personal profile designated by the consumer and made available to the company through a variety of means such as personal privacy profiles in conjunction with a clearinghouse (like the Do-Not-Call Registry and the Network Advertising Initiative) and data tagging. Each of these alternatives allows individuals to exercise individual control and obtain the benefits of transparency in instances in which often there is no relationship between the company and the individuals and it cannot be anticipated the ways in which data will be used.

Personal Privacy Profiles

The PCAST Report recommends the creation of standard sets of privacy preference profiles (that is, settings or choices) voluntarily offered by third parties (e.g. American Civil Liberties Union, Consumer Reports, Apple App Store, Google Play, Microsoft Store, financial services industry).[13] In the first instance, the organization offering the profile would vet new apps as acceptable or not acceptable within each of their profiles.[14]

CUSP supports the idea of personal privacy profiles but thinks they would be less burdensome to consumers if individuals only had to fill out one personal privacy profile and file it in one central location – a clearinghouse, perhaps one operated by a consumer protection agency within the federal government. Companies would then have the responsibility of checking this clearinghouse to ascertain whether they could make use of the data consistent with the expressed preferences of the consumer.

A clearinghouse of this sort currently exists at the Federal Trade Commission (FTC) in the form of the Do-Not-Call Registry. In 2003 the FTC amended the Telemarketing Sales Rule (TSR) to create the Do-Not-Call Registry which now includes more than 221 million unique telephone numbers.[15] The TSR prohibits telemarketers from calling an individual whose number is listed with the Do-Not-Call Registry.[16] By placing their numbers on the Do-Not-Call Registry, individuals are relieved of the burden of opting out of telemarketing by phone at each individual company.

The Do-Not-Call Registry has been enormously successful. The federal government should emulate it by creating a clearinghouse for personal privacy profiles where companies would go to check all the preferences of individuals.

---

[13] PCAST Report at 40.
[14] *Id.* at 41.
[15] *FTC 2014 Privacy and Data Security Update* available at http:// www.ftc.gov/reports/privacy-data-security-update-2014.
[16] *Id.*

Data Tagging

"Data tagging" refers to the inclusion of metadata with sufficient information that recipients and users understand their specific obligations for appropriate use and safeguarding of the underlying data.[17] Obligations to protect and secure metadata attach to the data itself and must be met wherever that data is stored or processed. Data protection obligations attach to and travel with the data providing individuals with protection in a decentralized, networked data environment – an environment where they currently have little knowledge of and little realistic choice about the actual party or parties handling their information.[18]

An example of "data tagging" is the Neptune and Cerberus pilots at the Department of Homeland Security.[19] For example, data attributes were identified and a set of tags to encode the information was developed. Tagging enabled both precise access control and preserved links to source data and the purpose of its original collection. The end result was a taxonomy of rules governing where information went and tracking where it came from and under what authority.[20]

The same concept could be applied to a personal privacy profile developed by the consumer. Once the consumer develops his or her profile, data associated with the individual is identified, tags are encoded with the consumer's privacy profile preferences, and the data is appropriately tagged. When a company obtains that data, the responsibility is on the company to use that data in accordance with the preferences expressed by the consumer. This way, consumers retain control over their data but do not have the burden of interacting with each company in possession of their data. The burden to act responsibly with respect to the use of the data falls appropriately on the entity possessing the data.

11. As the PCAST Report explains, "it is increasingly easy to defeat [de-identification of personal data] by the very techniques that are being developed for many legitimate applications of big data." However, de-identification may remain useful as an added safeguard in some contexts, particularly when employed in combination with policy safeguards. How significant are the privacy risks posed by re-identification of de-identified data? How can de-identification be used to mitigate privacy risks in light of the analytical capabilities of big data? Can particular policy safeguards bolster the effectiveness of de-identification? Does the relative efficacy of de-identification depend on whether it is applied to public or private data sets? Can differential privacy mitigate risks in some cases? What steps could the government or private sector take to expand the capabilities and practical application of these techniques?

Without getting into the debate regarding whether de-identification standing alone works or not,[21]CUSP agrees that de-identification remains useful as an added safeguard when employed with policy safeguards. When de-identification is used for internal, risk mitigation purposes, the risk of re-identification does not have to be non-existent for it to be effective. Risk management approaches used for years do not strive to obtain zero tolerance for risk; rather they identify risks, prioritize them and take appropriate steps to reduce them.

---

[17] Paula J. Bruening and K. Krasnow Waterman, *Data Tagging for New Information Governance Models*, IEEE Security & Privacy Magazine 8.5 (2010).
[18] *Id.*
[19] Big Data Report at 27.
[20] *Id.* at 28.
[21] *Compare* Ann Cavoukian and Daniel Castro, *Big Data and Innovation, Setting the Record Straight: De-Identification Does Work* available at http:www.privacybydesign.ca/index.php/peper/big-data-innovation-setting-record-straight-de-identification-work/ *with* Arvind Narayanan and Edward W. Felten, *No silver bullet: de-identification still doesn't work* at http://www.pogowasright.org.no-silver-bullet-de-identification-still-doesnt-work/.

This is the approach taken in the FTC's Standards for Safeguarding Customer Information issued under the Gramm-Leach-Bliley Act (Safeguards Rule).[22] Under the Safeguards Rule, financial institutions are required to identify reasonably foreseeable internal and external risks to the security, confidentiality and integrity of customer information that could result in the unauthorized disclosure, misuse, alteration, destruction or other compromise of that information and assess the sufficiency of any safeguards in place to control those risks. At a minimum, such a risk assessment should include consideration of risks in each relevant area of the financial institution's operations. The financial institution then should design and implement information safeguards to control the risks identified through the risk assessment and regularly test or otherwise monitor the effectiveness of the safeguards' key controls, systems and procedures. Also, the financial institution should evaluate and adjust its information and security program in light of the results of its testing and monitoring, any material changes to its operations or business arrangements, or any other circumstances that are known or that there is reason to know may have a material impact on its information security program.[23]

When used for internal purposes, de-identification is one of many tools in the risk reduction arsenal; it is one of the information safeguards that are implemented to control risks identified through risk assessments and regular testing and monitoring. De-identification is particularly effective when it is combined with policies prohibiting the re-identification of data that have clear consequences for prohibited re-identification, such as dismissal or suspension.

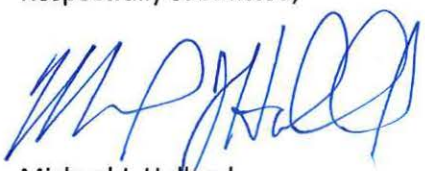
13. Can accountability mechanisms play a useful role in promoting socially beneficial uses of big data while safeguarding privacy? Should ethics boards, privacy advisory committees, consumer advisory boards, or Institutional Review Boards (IRBs) be consulted when practical limits frustrate transparency and individuals' control over their personal information? How could such entities be structured? How might they be useful in the commercial context? Can privacy impact assessments and third-party audits complement the work of such entities? What kinds of parameters would be valuable for different kinds of big data analysts to consider, and what kinds of incentives might be most effective in promoting their consideration?

Accountability mechanisms, which are the foundation of a responsible use framework, can play a useful role in promoting socially beneficial uses of big data while safeguarding privacy. See pages 2-4 of CUSP response to the "Big Data RFI."

Ethics boards, privacy advisory committees, consumer advisory boards, or Institutional Review Boards (IRBs) should be consulted when practical limits frustrate transparency and individuals' control over their personal information. IRBs should be able to allow reuse of data for research or education purposes. See pages 6 – 7 of CUSP response to the "Big Data RFI."

CUSP appreciates the opportunity to submit these comments and hopes they contribute to the Administration's review of the privacy issues raised by Big Data.

Respectfully submitted,



Michael J. Holland
Chief of Staff

[22] 16 CFR Part 314.
[23] *Id.*