

Big Data and Consumer Privacy in the Internet Economy

1

White House event attended businesses and advocacy groups and welcomed as striking a balance between innovation and growth and protecting consumers.

The Consumer Privacy Bill of Rights articulated in the White House privacy blueprint is based on Fair Information Practice Principles that have enduring value. They are the basis privacy protections incorporated into law and embedded in the privacy practices of businesses, hospitals, universities, and other institutions. They are the basis of global interoperability principles under the US-EU Safe Harbor Framework and the APEC Cross-Border Privacy Rules. But the 2012 blueprint broke new ground in recommending that these fundamental privacy principles must be “adapt[ed] to the dynamic environment of the commercial Internet” and therefore “flexible” and “dynamic.”

Consistent with this approach and as discussed further below, therefore, legislation should be principles-based and avoid detailed prescription, leaving specific application of the principles to evolving technology and practices, codes of conduct developed through multistakeholder processes advanced by legislation, and through case-by-case adjudication by the Federal Trade Commission. It is this baseline, adaptable approach that will enable the Consumer Privacy Bill of Rights to, in the words of the Request for Public Comment, “support the innovations of big data while at the same time responding to its risks” and lead to a “responsible use framework.”

I. Moving Forward With Legislation Will Reaffirm American Privacy Values

Enacting consumer data privacy legislation was one of the key elements of the 2012 White House privacy blueprint. That document, “urge[d] Congress to pass legislation adopting the Consumer Privacy Bill of Rights” by codifying the rights defined in the blueprint, and President Obama promised in the forward to “work with Congress to put [these principles] into law.” Assistant Secretary Larry Strickling and I as General Counsel both gave cleared testimony to congressional committees reiterating this position. Both the Fact Sheet accompanying the President’s January 17, 2014 speech on surveillance and the Big Data Task Force report reaffirmed the privacy blueprint.

Issuing proposed legislation will make good on the Administration’s commitment and provide a concrete roadmap for what the President called “a dynamic model of how to offer strong privacy protection and enable ongoing innovation in new information technologies.” These concurrent policy goals are also central to the Big Data Task Force Report and the report of the President’s Council of Advisors on Science & Technology. As discussed in greater depth below, the Administration’s model offers a novel approach intended to be as iterative and adaptive as the technologies and innovations at issue that departs from prescriptive regulation. An Administration bill will help stakeholders understand this approach.

The case for enacting consumer privacy legislation is even stronger today than in 2012. Both Big Data Task Force and PCAST reports documented the “explosion” in the volume, velocity, and variety of data that is becoming “near-ubiquitous” and “increasingly approaching real time” collection. The American system of privacy has great strengths: as described in those reports and the 2012 Blueprint, it rests on strong values deeply embedded in our culture and Constitution, sectoral regimes that protect the most sensitive categories of personal information, 47 state data breach notification laws and other laws, and active enforcement by the Federal Trade Commission and state attorneys general.

Privacy scholars Kenneth Bamberger and Deirdre Mulligan have documented the culture of active compliance led by privacy professionals that this system has fostered.⁵

Nevertheless, as the digital economy expands, an increasing proportion of data collection and use falls outside the sectors that are covered by privacy rules. One of the products of the big data reviews is a much deeper understanding of the extent to which rich data that can yield information every bit as sensitive as what is contained in health or financial records. Yet the increasing volume, velocity, and variety of data collection have made it impossible for individuals to exercise any meaningful control over most data about them. This results in a significant asymmetry in the marketplace. As the Big Data Task Force put it, “[unprecedented computational power and sophistication], most of which are not visible to the consumer, also create an asymmetry of power between those who hold the data and those who intentionally or inadvertently supply it.” This, by any definition, is a market failure.

This failure can produce two outcomes. First, it can harm individuals through abuses by government or private actors that are not responsible stewards of privacy. (Even when the U.S. Government and many U.S. companies are responsible stewards of data, some other governments and all data thieves are not). Second, it can erode the individual trust that is a necessary predicate to many of the benefits from sharing of data.

Moreover, in the wake of the Snowden revelations, there is an especially acute need to restore global trust in America’s protection of privacy and the companies that operate under American privacy law. As discussed in the White House big data review, data can be a public resource that can produce great benefits in medical research, urban planning, and public health among other areas. Establishing a digital environment in which people can trust that their data will be used in ways that are consistent with their interests and expectations is a critical enabler of such benefits.

Moving forward with a law that fills the gaps in U.S. protection of privacy would provide a strong reaffirmation of the Administration’s commitment to privacy. At a time when the European Union is move to adopt a new privacy regulation that would bind all member states, an Administration proposal would by presenting to the world a different, American model. It is an opportunity to seize leadership in the global discussion of privacy.

II. Proposed Legislation Should Retain a Principles-Based Approach.

A challenge in putting the Consumer Privacy Bill of Rights into legislative language is trying to anticipate all the questions that may arise and all the fact situations to which it may apply. There are trade-offs between certainty and creativity, between precision and flexibility, and the Department of Commerce should err on the side of uncertainty. The alternative is prescription. Indeed, the view adopted by the PCAST that “policies and regulation ... should not embed particular technological solutions, but rather should be stated in terms of intended outcomes” was a guiding principle in developing the Consumer Privacy Bill of Rights.

⁵ K.S. Bamberger & D.K. Mulligan, *Privacy in Europe: Initial Data on Governance Choices and Corporate Practices*, [George Washington Law Review](#), Vol. 81, p. 1529, [UC Berkeley Public Law Research Paper No. 2328877 \(2013\)](#); *New Governance, Chief Privacy Officers, and the Corporate Management of Information Privacy in the United States: An Initial Inquiry*, [Law & Policy](#), Vol. 33, Issue 4, pp. 477-508 (2011); *Privacy on the Books and on the Ground*, 3 Stan.L. Rev.247, 253 (2011).

The goal instead should be to *establish a framework* by which these myriad questions can be addressed. There are two mechanisms for doing so. The first is the multistakeholder process described in the White House blueprint and subsequently implemented by NTIA. This process allows for gathering information about technology or practices relevant to specific issues, as well as for adapting codes of conduct in response to changing issues.

The second mechanism is FTC adjudication of individual cases of violations of the Consumer Privacy Bill of Rights. This enforcement is based on the FTC's Section 5 authority where it finds unfair or deceptive practices or acts. Giving the FTC the direct authority under Section 5 to enforce the Consumer Privacy Bill of Rights would enable the FTC to enforce with respect to privacy practices without necessarily having to find deception or unfairness.

The FTC's current application of Section 5 to privacy and security practices operates as form of common law,⁶ feeling the stones along the bottom of the river as it iterates a growing body of law. That is the essence common law adjudication. The common law has applied uncertain constructs like "reasonable care" for centuries, and the common law delineated the American law of privacy protection from the 18th and 19th century cases described by Samuel Warren and Louis Brandeis until the adoption of the Privacy Act and Fair Credit Reporting Act in 1974. Federal common law has filled in the interstices of statutes such as antitrust and environmental laws. Under the expert hand of the FTC invigorated by authority to adjudicate the Consumer Privacy Bill of Rights, the common law is capable of doing the same for privacy in the Information Age.

For the same reasons as legislation should not venture into detailed prescription of privacy practices to implement the Consumer Privacy Bill of Right, it should not delegate such prescription to the Federal Trade Commission. The FTC does an exceptional job as the world's leading privacy enforcement agency, protecting not only Americans but – through the Safe Harbor Framework and consent decrees with Google, Facebook, and others – 100,000s of Europeans as well. It is capable of administering the Consumer Privacy Bill of Rights without substantive rulemaking authority.

Similarly, a number of the "specific questions" in the Request (such as "policy and technical solutions" for deletion of distributed information) are too granular to require answers as a predicate to drafting and proposing legislation. These are issues best left to code of conduct development with stakeholders in which there is an opportunity to explore technical issues and solutions and work out a response that can be adapted over time.

Accordingly, the legislative language should hew closely to the broad principles articulated in the 2012 White House policy blueprint. The issues are too complex, the technology too fast-moving, and the fact-situations too many to attempt addressing them specifically in legislative language.

III. Flexible Application of The Consumer Privacy Bill of Rights Fits the Age of Big Data.

The Request for Comment asks:

⁶ See D. J. Solove & W. Hartzog, *The FTC and The Common Law of Privacy*, [114 Columbia Law Review 583 \(2014\)](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2312913), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2312913; Mulligan & Bamberger, *supra*, *Privacy on The Books and on The Ground*.

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?

The Consumer Privacy Bill of Rights was conceived with the emerging world of big data in mind. Many of the numerous stakeholders who commented in NTIA's initial inquiry or on the Commerce green paper described emerging uses of data.⁷ Although when the Consumer Privacy Bill of Rights was released "big data" was not the everyday term it has become since, mobile apps and location services, social networking, online tracking, data mining, and multifaceted data collection such as Google's were at the core of the discussion.

The White House blueprint is premised explicitly on the increasing scale and scope of data collection: the President's forward refers to "an explosion of commerce and innovation ... enabled by novel uses of personal information," and the second paragraph of the introduction describes the elements of the big data explosion – "[a]n abundance of data, inexpensive processing power, and increasingly sophisticated analytical techniques ..." – and goes on to describe a number of big data applications in politics, public safety, media, and health.

With this emerging environment in mind, a key feature of the Consumer Privacy Bill of Rights is the flexibility most apparent in the discussion of the two Bill of Rights Principles that most change existing FIPPs: Respect for Context and Focused Collection. These demonstrate that the Consumer Privacy Bill of Rights is intended not as a formalistic checklist but as a set of principles that work together to foster an ecology of trust. It is precisely its holistic and interactive approach to the privacy principles rather than reliance on the single formalistic mechanism of notice-and-choice that makes the Consumer Privacy Bill of Rights "dynamic." In this light, it shares the goal of the PCAST to rest "the responsibility for using personal data in accordance with the user's preferences ... with the provider rather than with the user."

The Respect for Context principle ("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") has origins in the FIPPs principle of purpose specification; the latter calls for articulating the purpose(s) for which data is collected, which in turn operate as use limitations. Respect for Context is more flexible, however, because it does not necessarily tether use to purpose specification so long as another use is consistent with uses a consumer would expect based on either purpose specification or the relationship between consumer and company. (See pp. 18-19 "Respect for Context may not require a company to specify each use a recipient might make ..."). This flexibility was intended to allow for evolving uses, recognizing that in the face of rapidly changing uses of data, it is difficult to know today what new uses may develop 15 months from now. Thus, a company "does not need to seek affirmative consent each time it uses existing data to improve a new service, or even creates a new service, provided that these new uses of data are consistent with what users come to expect ... [in the circumstances of the example]" (Page 8).

⁷ See, e.g., Comments of Intel Corporation (Jan. 16, 2011) [<http://www.ntia.doc.gov/files/ntia/comments/101214614-0614-01/attachments/Intel%20Corp%20Dept%20Commerce%20green%20paper%20comment.pdf>].

The Focused Collection principle (“consumers have a right to reasonable limits in the personal data that companies collect and retain”) is even more specifically aimed at accommodating big data. Here, “reasonable limits” replace “data minimization” narrowly focused on data deemed necessary and relevant. This enlargement recognizes that hard-and-fast data minimization would unduly restrict the increasing use of unstructured databases, mining for unforeseen correlations, and other evolving uses of data. The Focused Collection principle acknowledges that “wide-ranging data collection may be essential for some familiar and socially beneficial Internet services and applications.” It therefore allows for evolving uses while calling for thoughtful and discriminating collection and retention of data.

This recognition that innovation needs a wide range of uses and requires a focus on context informs all of the Bill of Rights principles and underlies the flexibility of the blueprint. Legislative language therefore should make this flexibility clear, making clear as a general matter that a failure to implement a particular principle is not necessarily a violation of the Bill of Rights depending on the context and the other principles implemented to protect privacy. In addition to this essential interpretive guide, I offer the additional suggestions below with respect to specific provisions.

A. The Definition of Personal Data Should Be Broad.

The White House privacy blueprint used the term “personal data” and defined it as any data, including the aggregations of data, which is linkable to a specific individual. Although noting the similarity to widely-used term “personally identifiable information,” the blueprint specifically the broader, more flexible term.

The White House big data inquiry shows the appropriateness of doing so. It reflects the expanding science of re-identification techniques and increased understanding of the rich information about individuals that can be derived from metadata and other seemingly innocuous data. In this light, careful stewardship is needed for more than just fixed categories of personally identifiable information. Accordingly, the Consumer Privacy Bill of Rights should apply broadly.

B. The Individual Control Provisions Should Clarify That Consent Is Not Required in All Circumstances.

The Individual Control principle departs from the formalistic mechanism of notice-and-choice that has been widely and appropriately criticized not only by the PCAST, but also by the FTC and myself and other authors of the Consumer Privacy Bill of Rights. Instead, the blueprint recognizes that control should “reflect the scale, scope, and sensitivity of the personal data [companies] collect, use, or disclose as well as the sensitivity of the uses they make ...” and whether data is “reasonably linkable to individuals” Similarly, in certain circumstances, “[c]ompanies should be able to infer consumer consent”

The White House Task Force and PCAST reports correctly recognize that the increasing use of ubiquitous and automated sensing and other data collection makes notice-and-choice infeasible in some circumstances. Furthermore, used in all circumstances notice-and-choice as classically conceived – an explicit request (form, pop-up, etc.) accompanied by affirmative assent (a signature, a checked box) – devalues privacy because it becomes a rote process in which people check reflexively other circumstances where an individual has an opportunity to provide meaningful input and data is sensitive, consent still has a significant role to play (“remains fundamental in many contexts,” according

to the Task Force). The HIPAA Privacy Rule, for example, appropriately requires explicit consent for the disclosure of covered health information. Thus, the blueprint reflects that where sensitive information of this sort is involved, choice mechanisms should be “simple and prominent.” In a responsible use framework, mechanisms of control need to be tailored to individual circumstances.

And just because notice-and-choice does not work in some circumstances does not mean it is impossible to enable other forms of individual control. This is where other Bill of Rights principles can come into play; thus, the blueprint notes that “technology can help to expand the range of user control” through tools such as individual privacy setting, and “[i]f it is impractical to provide Individual Control, these companies should ensure that they implement other elements of the Consumer Privacy Bill of Rights in ways that adequately protect consumers’ privacy,” pointing to Transparency, Access and Accuracy, and Accountability.

The latter language should be incorporated into legislation, making clear that consent or notice-and-choice is not a requirement in all contexts and that alternative mechanisms of control based on other principles can satisfy this principle in appropriate contexts.

C. The Transparency Provision Should Not Be Tethered Exclusively to Privacy Policies.

Privacy policies are a necessary but not sufficient element of privacy protection. They serve an important role in articulating privacy practices and providing measuring sticks for regulators, journalists, privacy advocates, and vigilant consumers. They also operate to focus the thinking of the institutions. Thus, entities should be required to have privacy policies in place.

Nevertheless, research submitted to NTIA in its privacy inquiry shows that “[w]hen consumers see the term ‘privacy policy,’ they believe that their personal information will be protected in specific ways” that often are not the case.⁸ The remedy is not more or better privacy policies. Rather, the legislation should encourage development of additional mechanisms of transparency in most instances. These can include the “just in time” privacy notices, vetting by third-party organizations mentioned by the Task Force and PCAST, and use of Institutional Review Boards among other practices, with implementation driven by multistakeholder processes. They also can include mechanisms such as those adopted by data brokers to provide consumers access to information collected about and opportunities to correct the data (Axciom’s aboutthedata.com is an example). To alter the asymmetry of information the marketplace for data, the focus of transparency needs to shift from words in a privacy policy to tools that supply consumers with pictures of what data being collected and what that data says about them, and enable them to get greater benefit from data they themselves generate.

D. The Respect for Context Provision Should Consider Benefits As Well As Risks,

The Respect for Context principle seeks to understand consumers’ reasonable expectations in the circumstances based on the nature of the data collected and the situation in which it is collected. As discussed above, it is a linchpin of the Consumer Privacy Bill of Rights, affecting the application of all other principles.

⁸ J. Turow, C.J.Hoofnagle, D.K.Mulligan,N.Good & J. Grossklag, *The Federal Trade Commission and Consumer Privacy in the Coming Decade*, 3 I/S: J. of Law & Policy for the Info. Soc. 733, 734 (2007), www.ntia.doc.gov/files/ntia/comments/100402174-0175-01/attachments/FTC_and_privacy.pdf.

An important factor to consider in measuring data collection, use, or disclosure according to context is the potential benefit to the consumer in addition to the risk. Just as the risk of some harm to the consumer is an important factor – consumers objectively do not expect their data to be used against their interest – so is the use of data in ways that benefit their interests. Medical researchers at the MIT Big Data Workshop in March, for example, pointed out that most medical patients are happy to have their medical records used to find cures for their ailments even if the cure will not help them. Use of data from traffic cameras or sensors to speed the flow of traffic presents different issues than use of the same data for surveillance of populations. The legislative discussion of context should make explicit the impact of risk and benefit, or beneficial and adverse interests.

The PCAST is correct that the language referring to context in terms how “consumers provide the data” may not adequately address circumstances where data is collected without any involvement by the consumer. The language could be enlarged or changed to refer to the context in which the data is “collected.” The Respect for Context language used in the Consumer Privacy Bill of Rights focuses on the relationship between the consumer and company or companies collecting, using, or disclosing the consumer’s data; the absence of any relationship is one context that implies limited uses of the consumer’s data.

The potential for invidious discrimination or other adverse uses of data are significant risks to be considered. The decision whether certain inferences involve discrimination that is invidious is fundamentally a matter of social and human rights policy (racial discrimination a stark example; the Genetic Information Nondiscrimination Act and state laws prohibiting employers from requiring applicants to supply social media passwords are others more directly linked to data use). Privacy law cannot carry the full freight for such policies, and thus it is not necessary in the context of privacy legislation to address potentially discriminatory uses. But the Consumer Privacy Bill of Rights can help by promoting greater transparency, accountability, and thoughtfulness in how data is used, making it easier to identify and to prevent unwanted inferences.

An approach to context that balances risk and benefit to individuals will help to promote a responsible use framework. According to the blueprint, “The Respect for Context Principle calls on companies that collect data to act as stewards of data in ways that respect their consumers.” In a sense, it brings privacy protection full circle. When Samuel Warren and Louis Brandeis wrote their seminal Right to Privacy article, they inferred the right from the common law. One of the main lines of authority they drew on was the law of implied trust – the principle that one entrusted with confidential information owes a duty to the person whose information it is. Trust law separates legal ownership, custody, and control from the benefit from of assets, and imposes on the trustee duties to protect the interests of the beneficial owner and avoid self-dealing. A good steward does the same.

E. Legislation Should Include a Provision Addressing Re-Identification.

One of the issues examined in the White House Big Data review both by the task force and by the PCAST was de-identification and re-identification, the technologies involved, and the growing challenges they present for beneficial research. The task force noted the “view that de-identification of data as a means of protecting individual privacy is, at best, a limited proposition” and corresponding concerns that “meaningful de-identification may strip the data of both its usefulness and the ability to

ensure provenance and accountability.” The Request for Comment raises questions about re-identification based on the PCAST report.

There is debate among privacy experts about the effectiveness of de-identification (or anonymization) in light of well-known demonstrations of re-identification by computer scientists and journalists. This debate was renewed recently, with defenders of de-identification⁹ arguing demonstrations of re-identification overstate the risk and skeptics responding¹⁰ that the increasing array of data from which to connect anonymous data with known individuals and the increasing sophistication of techniques to identify unique cases have made it increasingly possible to identify individuals even from data anonymized with care.

De-identification has a role to play in a dynamic privacy framework as one of numerous privacy practices that can help protect privacy of data after it is collected. Few defenders of anonymization argue that it provides sufficient privacy protection by itself. And many critics concede anonymization has some place in the privacy toolbox so long as it is used thoughtfully alongside other tools of privacy protection.

Public policy can provide a layer of stronger assurance that anonymous data will stay anonymous. We therefore recommend that legislation include a provision creating civil and criminal liability for re-identification of data that has been intentionally de-identified. As the PCAST wrote in its report, “measures are most effective when there are regulations and laws with civil and criminal liability.” Such a provision would express our privacy values and shift for the better the risk-benefit balance in the use of data. The norms of behavior established thereby would influence individual and institutional practices and cultures, and reduce the number of people who might undertake re-identification and to reduce the availability of data that in the aggregate can be used for re-identification.

Such a provision should allow some exceptions where access to identity is needed for socially-valuable purposes. These would include: (1) duly-authorized government access (2) research into re-identification methods treated and approved as human experimentation, which helps in understanding privacy risks and promotes stronger privacy protection; (3) emergencies of public safety or public health; and (closely related to the latter) (4) where there is reason to believe the health or safety of the person(s) identified is sure they accomplish what is needed without creating loopholes. Re-identification also would present significant First Amendment issues in some circumstances, but these are better adjudicated case-by-case in as-applied challenges as they arise.

The principal objection to such a provision is the question of enforceability. Technology and legal measures can help, such as watermarking, chain-of-custody systems, and contractual arrangements for example. Moreover, some significant categories of data users are likely to comply with a legal regime on re-identification. These include academic researchers, companies whose privacy policies require them to strip identifiers from data they collect, and government agencies. Such entities in varying degrees lack motivation to re-identify and would face the additional deterrent of

⁹ A. Cavoukian & D. Castro, *Big Data And Innovation, Setting The Record Straight: De-identification Does Work* (June 16, 2014), <http://www.privacybydesign.ca/index.php/paper/big-data-innovation-setting-record-straight-de-identification-work/>.

¹⁰ A. Narayan & E. Felten, *No Silver Bullet: De-Identification Still Doesn't Work* (July 9, 2014), <http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf>.

violating law as well as institutional policies. In addition, the prospect of lawsuits or criminal enforcement would get the attention of corporate c-suites, counsel, and compliance officers and bring the protection of privacy and identity into evaluation of information security practices and risks. Legal sanctions thus would enable greater sharing of data in the areas where society stands to benefit the most.

Conclusion

Submission of an Administration bill to Congress on the Consumer Privacy Bill of Rights will continue the “national conversation on privacy” in which President Obama has been engaged the past year and initiate a robust discussion of the Administration’s innovative regulatory model for consumer privacy. This model strikes a good balance and offers a path forward to meet the challenges of privacy in the age of big data.

Respectfully submitted,

/s/

Cameron F. Kerry

**Ann R. and Andrew H. Tisch Distinguished
Visiting Fellow**

Governance Studies and Center for Technology &
Innovation

The Brookings Institution

1755 Massachusetts Avenue, NW

Washington, DC 20036-2103

ckerry@brookings.edu

www.brookings.edu/experts/ckerry

202-797-6090 (Governance Studies)

Visiting Scholar

MIT Media Lab

Room E15-384a, 30 Ames Street

Cambridge, MA 02139

ckerry@media.mit.edu

617-253-5667

Twitter: @Cam_Kerry