August 5, 2014

Via e-mail: privacyrfc2014@ntia.doc.gov

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

Re: Privacy RFC 2014

The Future of Privacy Forum (FPF) is a think tank seeking to advance responsible data practices and includes leaders in business, academia, and consumer advocacy. FPF appreciates the opportunity to provide these Comments in response to NTIA's June 5, 2014 Request for Comment (RFC), seeking public input on how developments related to "big data" impact the Consumer Privacy Bill of Rights.[1]

Unlocking the value of data and instituting responsible data practices go hand-in-hand, and both have been an important focus of FPF's work since our founding in 2008. FPF recognizes the enormous potential benefits to consumers and to society from big data analytics,[2] and FPF also understands that taking advantage of big data will require traditional privacy principles to evolve. The Consumer Privacy Bill of Rights likewise endorses a flexible approach to the use of different practices and tools to protect privacy. With respect to big data, flexibility in the application of privacy protections is essential.

FPF supports efforts by the NTIA to consider how the Consumer Privacy Bill of Rights can support uses of big data, and we have previously recognized the need for further conversations about how privacy can be protected in the age of big data.[3] Turning to specific questions posed by the RFC:

(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 posted by big data?

The Consumer Privacy Bill of Rights recognizes that protecting privacy requires implementing practices that are "comprehensive, actionable, and flexible."[4] The general principles put forward by

[1] NTIA, Big Data and Consumer Privacy in the Internet Economy, Request for Public Comment, 79 Fed. Reg. 32,714 (June 5, 2014).
[2] Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 243-51 (2013).
[3] In September 2013, FPF and the Stanford Center for Internet & Society held a day long symposium exploring how to address privacy issues raised by big data. A collection of papers prepared in advance of the event is available at http://www.futureofprivacy.org/big-data-privacy-workshop-paper-collection/.

the Administration's privacy framework explicitly afford companies discretion in how they implement them. This flexibility was designed both to promote innovation and to "encourage effective privacy protections by allowing companies, informed by input from consumers and other stakeholders, to address the privacy issues that are likely to be most important to their customers and users, rather than requiring companies to adhere to a single, rigid set of requirements."[5] For example, while the framework calls for offering individual control of information, it recognizes that this may be impractical in some cases. Instead, organizations are encouraged to embrace other elements of the Consumer Privacy Bill of Rights and augment internal practices in order to adequately protect consumer privacy.[6]

Such flexibility can be accommodated within the established framework provided by the Fair Information Practice Principles (FIPPs). Data innovation requires that the FIPPs are not inflexible "one size fits all" rules but rather must be adapted to the circumstances of data collection and use.[7] For example, use limitation is a long-standing, valuable FIPP, but it is strained in an age of big data. If use limitation is delimited by the purposes specified at the time of collection, we may never be able to obtain the unexpected benefits that are promised by big data. As the Consumer Privacy Bill of Rights recognizes, privacy is promoted when data are used in ways that respect the context of collection. Context changes over time. Respecting the context of collection must allow room for innovative uses of data that deliver unexpected benefits. And respect for context must be framed in a way that reflects the dynamic nature of social and cultural norms and the subjective nature of consumer trust.

FPF submits that the Consumer Privacy Bill of Rights should be clarified to make explicit the need for flexibility with respect to big data. In particular, and as discussed below and endorsed by the recent White House Big Data Report[8], big data warrants a shift toward greater emphasis on responsible use. A use-based approach is one that respects the *context* in which information is collected and used, and implements stronger accountability and enforcement mechanisms. As discussed further below, the context principle should encompass factors such as the type of data being used and how the data are being used, including the broader societal benefits that could result and consideration of relevant risks.

Pragmatic contextual considerations can also inform how other existing privacy principles, including notice, choice, and data minimization, are applied in practice. Again, flexibility will be key, and the principles within the Consumer Privacy Bill of Rights should be properly calibrated both to protect privacy and encourage innovative uses of data. Longstanding practices emphasizing limits on collection and use impose rigid rules of the road that dismiss the flexibility inherent in calls for increased transparency and individual control. A flexible approach will allow organizations to

---

[4] WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY 9 (2012), http://www.whitehouse.gov/sites/default/files/privacy-final.pdf [hereinafter WHITE HOUSE BLUEPRINT].
[5] *Id.* at 2.
[6] *Id.*at 9.
[7] *See, e.g.*, Comments of the Future of Privacy Forum RE: Internet of Things, Project No. P135405, at 3-5 (Jan. 2014), *available at* http://www.futureofprivacy.org/wp-content/uploads/FPF-IoT-Comments_January-2014.pdf.
[8] WHITE HOUSE EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 56 (May 2014), http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [hereinafter BIG DATA REPORT].

implement benefit-risk analyses and accountability mechanisms that will be conducive to a considered approach to privacy and big data.

At the same time, big data also highlights the importance of data security. One recent survey suggested that 80% of consumers are more worried that information they share will be "hacked" or stolen, while 16% were more concerned that their information will be shared for commercial purposes or used to target advertising to them.[9] According to the Federal Trade Commission, identity theft is regularly the top consumer worry,[10] and FTC Chairwoman Edith Ramirez has cautioned that big data will exacerbate the threat posed by data breaches.[11] The Consumer Privacy Bill of Rights recognizes the importance of security, but it must remain a top priority moving forward.

(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?

New technologies, including big data and the emerging Internet of Things, are increasingly better suited to a responsible use framework.[12] In particular, one of biggest challenges – and benefits – posed by big data is that much of the new value from data is being discovered in surprising ways.[13] Accordingly, it is not always possible to provide precise notice in advance of collection on subsequent use of data. Calls for a responsible use framework are a response to concerns that traditional applications of existing privacy principles both constrain valuable uses of data and do not accurately reflect how consumers use new technologies.[14]

A responsible use framework is reflected throughout the Consumer Privacy Bill of Rights. Specifically, the Consumer Privacy Bill of Rights calls for respecting the context in which information is collected and used. The principle of respect for context makes companies stewards of consumer data.[15] Context reflects and respects how individuals think about their personal data,[16] and it provides a framework for organizations to ensure data use is in line with consumer expectations.

---

[9] Heather Greenfield, Computer & Communication Industry Association, *Major Study Sheds Light On Online Privacy, Security Values, Behavior* (Dec. 20, 2013), http://www.ccianet.org/2013/12/major-study-sheds-light-online-privacy-security-values-behavior/.

[10] Colleen Tressler, Fed. Trade Comm'n, *Identity Theft Tops List of Consumer Complaints for 14th Consecutive Year* (Feb. 27, 2014), http://www.consumer.ftc.gov/blog/identity-theft-tops-list-consumer-complaints-14th-consecutive-year.

[11] Edith Ramirez, Chairwoman, Fed. Trade Comm'n, Keynote Address at the Technology Policy Institute Aspen Forum, The Privacy Challenges of Big Data: A View from the Lifeguard's Chair (Aug. 19, 2013), *available at* www.ftc.gov/public-statements/2013/08/privacy-challenges-big-data-view-lifeguard's-chair.

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

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

[14] FRED CATE & VIKTOR MAYER-SCHÖNBERGER, DATA USE AND IMPACT GLOBAL WORKSHOP (Dec. 1, 2013), http://cacr.iu.edu/sites/cacr.iu.edu/files/Use_Workshop_Report.pdf.

[15] WHITE HOUSE BLUEPRINT, *supra* note 4, at 16.

[16] Christopher Wolf, The Role of Data Use Analysis in Measuring and Protecting Against Privacy Harms 5 (June 5, 2014) (unpublished manuscript, on file with Future of Privacy Forum).

When personal information is used in ways that individuals would reasonably expect, privacy concerns are avoided. Context can be broken down by examining: 1) the type of data being used; 2) the service relationship in which data are collected; 3) how the information is used; 4) the type of device on which the transaction is being made; 5) whether the data was collected passively or actively; and 6) the potential benefits and harms that could result.[17] In addition, the direct use of data on a first-party basis may raise different privacy considerations than data that is shared with third parties or disclosed publicly. The context principle rests on a number of subjective and fact-specific variables, including an individual's level of trust in an organization and his or her perceptions of the value he or she derives from the use of the information.[18] Organizations and policy makers will need to acknowledge and remain aware of ever-shifting social and cultural norms in evaluating contextual use.[19]

Respect for context is only part of a responsible use framework. The Big Data Report recognizes that contextual principles may complement emerging use-based approaches to thinking about data.[20] There are uses of data that are outside of the traditionally-understood bounds of context that may also have high societal value. Ideally, a use-based approach aspires to consensus around broadly acceptable data uses—or on data uses that are deemed illegitimate, allowing organizations and policy makers to focus on managing the risks associated with middle of the road cases.[21]

When data are used in ways that may be out of context, the results may still be acceptable – and beneficial to consumers. Increasingly, the data collected from a range of apps as well as wearable and other "smart" devices could be used in ways that are out of context but useful to consumers. For example, smart store technologies can leverage the interaction of store Wi-Fi networks and mobile phones. Stores can learn how long consumers are waiting in line to check out and can understand how many consumers are repeat shoppers or which window displays are successful at bringing consumers into the store or to a register.[22] Increasingly, retailers hope to use real-time analytics data to better serve customers, make the shopping experience more enjoyable, and better compete with the convenience of simply shopping online.[23] FPF worked with a number of stakeholders to ensure that smart store technologies can be deployed in a way that protects consumer privacy and promote responsible uses of this information.[24]

Moving forward, when data are used out of context or otherwise in a manner that challenges the principles in the Consumer Privacy Bill of Rights, organizations should demonstrate that their big

---

[17] *See id.*

[18] Carolyn Nguyen, Director, Microsoft Technology Policy Group, Contextual Privacy, Address at the FTC Internet of Things Workshop (Nov. 19, 2013) (transcript available at http://www.ftc.gov/sites/default/files/documents/public_events/internet-things-privacy-security-connected-world/final_transcript.pdf.).

[19] *Id.*

[20] The White House Big Data Report recognizes this tension. *See* BIG DATA REPORT, *supra* note 8, at 56.

[21] SCOTT CHARNEY, MICROSOFT, TRUSTWORTHY COMPUTING NEXT 22 (2012); *see also* Scott Charney, *The Evolving Pursuit of Privacy*, HUFFINGTON POST (Apr. 10, 2014 3:04 PM), http://www.huffingtonpost.com/scott-charney/the-evolving-pursuit-of-p_b_5120518.html.

[22] *See* SmartStorePrivacy.com (last visited July 1, 2014).

[23] *See* Cindy Waxer, *Brick and Mortar Retailers Enlist Real-Time Analysis to Counter Online*, DATA INFORMED (Apr. 29, 2014), http://data-informed.com/brick-mortar-retailers-enlist-real-time-analysis-counter-online/.

[24] Press Release, The Future of Privacy Forum and Sen. Schumer Announce Important Agreement to Ensure Consumers Have Opportunity to "Opt-Out" Before Stores Can Track Their Movement Via Their Mobile Devices (Oct. 22, 2013), http://www.futureofprivacy.org/2013/10/22/schumer-and-tech-companies-announce-important-agreement-to-ensure-consumers-have-opportunity-to-opt-out-before-stores-can-track-their-movement-via-their-cell-phones/.

data projects have undergone a benefit-risk analysis that weighs the issues at hand. FPF believes that Chief Privacy Officers, privacy boards, and broader ethics panels can provide a practical outlet to preserve innovation and demonstrate accountability.

In addition to providing a foundation for responsible use, establishing effective enforcement and well-defined accountability mechanisms are also essential to increase the global interoperability of privacy laws.[25] The White House's "Privacy Blueprint," which includes the Consumer Privacy Bill of Rights," specifically flags the use of multi-stakeholder processes to develop voluntary but enforceable codes of conduct to implement its principles across a range of sectors.[26] FPF supports the continued development of these codes and the pathway to interoperability as described in the Administration's Blueprint. Global interoperability is particularly important in the context of big data.[27] If big data is balkanized into geographic regions, its value is diminished. Without a common understanding of what constitutes the responsible use of big data, conflicting approaches at the national level can and will take root, further diminishing the economic gains that can be realized through innovation. FPF encourages the promotion and maintenance of existing frameworks that promote interoperability. These frameworks include the U.S.-EU Safe Harbor and the APEC Cross Border Privacy Rules System. Additionally, FPF encourages the development and consideration of new frameworks, such as those that might emerge in ongoing trade negotiations.

(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?

Notice is often considered the most "fundamental" principle of privacy protection.[28] Yet there is wide acknowledgement that a privacy framework based solely on notice and choice has significant limitations.[29] The vast majority of consumers do not read privacy policies,[30] and further, studies have shown that consumers make privacy decisions not based on policies but rather on the context in which they are presented by a use of their data.[31] More detailed privacy policies should not be the sole solution, and policy makers should be wary of inundating users with more and more notices. Instead, in the age of big data, notice and choice mechanisms will need to be supplemented by additional tools to effectively protect privacy. As a practical matter, connected devices or other "smart" technologies that will not be equipped with interactive screens or other easily accessible user interfaces will prove challenging for traditional implementations of notice and choice. Information

[25] WHITE HOUSE BLUEPRINT, *supra* note 4, at 31.
[26] *Id.* at 25.
[27] The Big Data Report recognizes that "[t]he benefits of big data depend on the global free flow of information." BIG DATA REPORT, *supra* note 8, at 63.
[28] FEDERAL TRADE COMMISSION, PRIVACY ONLINE: A REPORT TO CONGRESS 7 (1998).
[29] Fred Cate, *Looking Beyond Notice and Choice*, PRIVACY & SECURITY LAW REPORT (Mar. 29, 2010), http://www.hunton.com/files/Publication/f69663d7-4348-4dac-b448-3b6c4687345e/Presentation/PublicationAttachment/dfdad615-e631-49c6-9499-ead6c2ada0c5/Looking_Beyond_Notice_and_Choice_3.10.pdf (citing Former FTC Chairman Jon Liebowitz conceding that the "notice and choice" regime offered by the FIPPs hasn't "worked quite as well as we would like.").
[30] Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J. L. & POL'Y FOR INFO. SOC'Y 543 (2008).
[31] *See, e.g.*, Alessandro Acquisti et al., What Is Privacy Worth? 27-28 (2010) (unpublished manuscript), *available at* http://www.heinz.cmu.edu/~acquisti/papers/acquisti-ISR-worth.pdf.

collected in "public" spaces and used for data analytics also suggest the need for novel approaches to privacy protection.

Flexibility will be especially important with respect to the concepts of notice and consent. Organizations need to think creatively about how to provide consumers with meaningful insight into commercial data practices when necessary, and regulators and policymakers should encourage these efforts.[32] Techniques to inform consumers of data practices might include symbols, short phrases, colors, diagrams, or any of the tools otherwise available to designers seeking to provide users with an engaging user experience. In some cases, just in time notices may be most effective. In others, it may be important to explain to consumers the benefit of the bargain in a document they can examine in advance. In the end, design features that "communicate" information to users may be more helpful than traditional notice models. The Consumer Bill of Rights should continue to express support for these innovative developments through flexible multi-stakeholder processes, as opposed to establishing overly restrictive rules that limit the collection and use of data.

Public-facing efforts to inform consumers about big data may improve awareness and understanding more than rigid applications of the notice and consent model. The NTIA should recognize and encourage efforts by companies to engage consumers in meaningful conversations where both parties' interests and expectations can be aligned. FPF has previously described the benefits of data "featurization," transforming data analysis into a consumer-side application by granting individual access to their personal data in intelligible, machine-readable forms.[33] Services such as personal clouds or data stores will allow individuals to contract with third-parties who would get permission to selectively access certain categories of their data to provide further analysis or value-added services.[34] "Featurization" could allow individuals to declare their own policies, preferences and terms of engagement, and do it in ways that can be automated both for them and for the companies they engage.[35]

11) 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?

---

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

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

[34] Tene & Polonetsky, *supra* note 2, at 263-70.

[35] The rise of privacy and reputation management vendors points to a future where organizations will be able to unlock additional value by collaborating with their users. One interesting first step is the launch of "About the Data" by Acxiom, the nation's largest data broker. "About the Data" is a consumer-facing tool that gives individuals control over certain categories of information (such as personal characteristics, interests, and finances) gathered by Acxiom. The site allows consumers to correct information, suppress any data they see, or opt-out of Acxiom's marketing profile system altogether via an approachable user interface.

As the President's Council of Advisors on Science and Technology Big Data Report recognizes, de-identification remains a useful tool to protect privacy in an age of big data.[36] However, the report remains considerably skeptical about the technical ability to ensure de-identified data cannot be re-identified.[37] These concerns result from the fact that researchers have routinely discovered ways to re-identify de-identified datasets that have been released publicly. But these examples often overstate the failure of de-identification.[38] De-identification must be assessed within the context of the type of data involved and how it is being used, and not the merely the mathematical possibility that data may be re-identified.[39]

Prominent examples of successful re-identification demonstrate the challenges of perfectly de-identifying datasets that are *released publicly*. Once data are made publicly available, there is an opportunity for any attacker with the time, resources, or technological capability can attempt to re-identify information. While perfect de-identification may not be technically achievable, the actual risk of re-identification of individuals from properly de-identified data is incredibly low.

Much of our discourse around de-identification focuses on the technical possibility of re-identification and the assumption that all data will be made publicly available.[40] This emphasis does not describe the entire universe of data that exists today. It is essential to differentiate the risks between privately-held and publicly-available information. Tying discussions of de-identification to its effectiveness with public datasets dismisses its value for safeguarding data that is either kept internally or only shared with a limited audience. To be sure, when de-identified data are disclosed publicly with only minimal technical controls under controlled conditions in place to mask identifiers, the risks of identification are much greater.

The successful re-identification of Netflix users is an example of how low the actual risk of re-identification is with respect to publicly released data.[41] In a recent study, researchers compared data widely released by Netflix with records available on the Internet Movie Database in order to uncover the identities of Netflix users. However, only two out of 480,189 Netflix users were successfully identified with this method. That equates to a 0.0004 percent rate of re-identification. That is a significantly lower risk that an individual's lifetime risk of being struck by lightning, which is approximately one in seven thousand.[42] Though one should not ignore the invasion of privacy

---

[36] PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE & TECH., BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE 38 (May 2014), http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

[37] *Id.*

[38] *E.g.* ANN CAVOUKIAN & DANIEL CASTRO, BIG DATA AND INNOVATION, SETTING THE RECORD STRAIGHT: DE-IDENTIFICATION DOES WORK (June 2014), http://www2.itif.org/2014-big-data-deidentification.pdf.

[39] Comments of the Future of Privacy Forum to the Federal Communications Commission, WC Docket No. 13-306 (Jan. 2014), *available at* http://www.futureofprivacy.org/wp-content/uploads/01-17-2014-FPF-Comments-to-the-FCC.pdf.

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

[41] *Id.* at 6 (citing Arvind Narayanan and Vitaly Shmatikov, Robust De-anonymization of Large Sparse Datasets, Proceedings of the 2008 IEEE Symposium on Security and Privacy (2008)).

[42] Joseph Jerome, *Making Perfect De-Identification the Enemy of Good De-Identification*, FUTURE OF PRIVACY FORUM (June 19, 2014), http://www.futureofprivacy.org/2014/06/19/making-perfect-de-identification-the-enemy-of-good-de-identification/.

experienced by these two individuals, the chief examples of de-identification's failings ultimately produced minimal privacy harms.

Moreover, the data released by Netflix could have been more effectively de-identified using methods such as the HIPAA de-identification Safe Harbor standards. In other words, the better conclusion to be drawn from the Netflix example is not that de-identification is ineffective, but rather that a more rigorous standard should have been applied to the dataset.[43]

De-identification must be assessed in context, and when appropriate technical standards are combined with reasonable legal and administrative safeguards, de-identification is an important tool to protect the privacy of individuals. Policy makers should recognize that de-identification techniques encompass a range of different solutions that can be deployed depending upon the potential threats at stake or proposed use of data.

For information that is not made public, practical de-identification can be supported by an array of safeguards including access controls and restrictions, contractual data use restrictions, and data deletion protocols that can augment technical measures in order to protect privacy. When these tools used in combination, a bad actor must circumvent administrative restraints *and* then re-identify any data before getting any value from his malfeasance – which, in practice, is difficult to accomplish.[44]

The Federal Trade Commission's reasonableness standard for de-identification is highly informed by non-technical factors, specifically the "nature of the data at issue and the purposes for which it will be used."[45] As the FTC suggests, the level of technical de-identification should take into account the sensitivity and potential value of the data at stake. The level of de-identification should be calibrated to consider the sensitivity of information, the level of identifiability of the data, the potential privacy risks at stake, *and* data utility for researchers and industry. The FTC standard also recognizes the importance of good internal processes and controls on the sharing of data with third parties. FPF believes that incorporating a benefit-risk assessment will help organizations tailor the level of de-identification necessary for particular uses of data.

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?

As discussed above, a responsible use framework will require organizations to develop new accountability mechanisms to ensure that they are managing personal information is responsible –

[43] CAVOUKIAN & CASTRO, *supra* note 38, at 6. Further, it is worth recognizing that the content of Netflix' records by themselves were not sufficient to re-identify anyone – it required access to information beyond the records themselves.
[44] *Id.*
[45] FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID Change 37 (Mar. 2012), http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

and ethical – ways. The NTIA should recognize the important role that internalized and formal review processes can have in considering innovative data projects and identifying any potential risks.[46] In addition, big data "algorithmists" or other professionals could also function inside organizations to evaluate the selection of data sources, the choice of analytical tools, and the interpretation of any predictive results.[47] As organizations increasingly face interesting new proposals for using data, these professionals could operate across the public and private sectors and conduct risk-benefit analyses of data uses.

Industry increasingly faces ethical considerations over how to minimize data risks while maximizing benefits to all parties.[48] Big data projects sometimes raise issues that transcend traditional privacy concerns and implicate broader policy concerns about new forms of discrimination or filter bubbles that could chill democratic discourse.[49] As the Big Data Report recognizes, there is a potential tension between socially beneficial and privacy invasive uses of information in everything from educational technology to consumer generated health data. The advent of big data requires active engagement by both internal and external stakeholders to increase transparency, accountability and trust.

A documented review process may serve as an effective tool to infuse ethical considerations into data analysis without requiring radical changes to the business practices or innovators or industry in general.[50] Institutional review boards (IRBs), which are the chief regulatory response to decades of questionable ethical decisions in the field of human subject testing, provide a useful precedent for focusing on good process controls as a way to address potential privacy concerns. While IRBs have become a rigid compliance device and would be inappropriate for wholesale use in big data decision-making,[51] they could provide a useful template for how projects can be evaluated based on prevailing community standards and subjective determinations of risks and benefits, particularly in cases involving greater privacy risks. Using an IRB model as inspiration, big data may warrant the creation of new advisory processes within organizations to more fully consider ethical questions posed by big data.

Traditional IRBs provide only one example of what this might look like, however. Privacy officers and other internal processes, such as privacy impact assessments and review boards, can provide an important outlet for evaluating certain categories of data uses. Documentation efforts could be particularly important when a proposed use might exceed the principles of the Consumer Privacy Bill of Rights. For example, a company may believe that keeping data beyond the minimum time needed may yield important benefits or may discover a use that was not specified in a notice. By carefully considering the potential risks and benefits, a company should be able to advance such a use and be consistent with any proposed law.

Moving forward, broader big data ethics panels could provide a commonsense response to public concerns about data misuse. While these institutions could provide a further expansion of the role of

---

[46] *See* Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 STAN. L. REV. ONLINE 97 (2013).
[47] SCHÖNBERGER & CUKIER, *supra* note 13, at 180.
[48] danah boyd, *What Does the Facebook Experiment Teach Us?*, MEDIUM (July 1, 2014), https://medium.com/message/what-does-the-facebook-experiment-teach-us-c858c08e287f.
[49] Jules Polonetsky & Omer Tene, *The Facebook Experiment: Gambling? In This Casino?*, RE/CODE (July 2, 2014 11:55 AM PDT), http://recode.net/2014/07/02/the-facebook-experiment-is-there-gambling-in-this-casino/.
[50] Matthew Salganik, *After the Facebook Emotional Contagion Experiment: A Proposal for a Positive Path Forward*, FREEDOM TO TINKER (July 7, 2014), https://freedom-to-tinker.com/blog/mjs3/after-the-facebook-emotional-contagion-experiment-a-proposal-for-a-positive-path-forward/.
[51] See Calo, *supra* note 46.

privacy professionals within organizations, they might also provide a forum for a diversity of viewpoints inside and out of organizations. Ethics reviews could include members with different backgrounds, training, and experience, and could seek input from outside actors including consumer groups and regulators. It is likely that these review boards will vary from organization to organization, and standardization would present a unique set of challenges, but organizations should be encouraged to develop structures and personnel to grapple with big data and provide an important check on any data misuse.

Any successful approach to big data must recognize the different ways that data can now be used. Many uses of big data are machine-to-machine or highly aggregated and do not implicate privacy concerns. It is also the case that many new uses of data are marginal and should not require enhanced review processes. Furthermore, the process of mitigating risks is often iterative and nuanced to reflect strategies that can well address minor changes to products or services.

That said, the more challenging areas of big data should be guided by a risk-benefit analysis that takes into account exactly how the benefits of big data will be distributed. So far, our procedural frameworks are largely focused on traditional privacy risks and assessing what measures can be taken to mitigate those risks. In 2010, for example, the Department of Commerce's Internet Policy Task Force endorsed the use of privacy impact assessments (PIAs) both to help organizations decide whether it is appropriate to engage in innovative data uses and to identify alternative approaches that could reduce relevant privacy risks.[52] However, human research IRBs also take into account anticipated benefits and even the importance of any knowledge that may result from research.[53]

Organizations and privacy professionals have become experienced at evaluating risk, but they should also engage in a rigorous data benefit analysis in conjunction with traditional privacy risks assessments. FPF suggests that organizations could develop procedures to assess the "raw value" of a data project, which would require organizations to identify the nature of a project, its potential beneficiaries, and the degree to which those beneficiaries would benefit from the project. This value would be discounted by the probability that a big data benefit can actually be accomplished. While there are no definitive rules of what probability of benefit is needed to overcome a presumption against exposing a beneficiary to privacy risk, it is clear that the mere assertion that a product or service can be improved is inadequate; yet demanding organizations provide proof beyond any doubt would be an impossible standard. Further, any analysis would need to take into account different community standards and expectations, as well as an inherent degree of subjectivity in assessing the relative merits of various projects. FPF's initial guidance on this type of analysis is attached as Appendix A.

However, this limitation has not stopped the federal government from developing structured processes to measure project benefits against risks. The White House Office of Management and Budget, for example, requires federal agencies to engage in comprehensive cost-benefit analysis. This analysis recognizes that even where values are not computable, "efforts to measure it can produce useful insights even when the monetary values of some benefits or costs cannot be determined."[54]

---

[52] U.S. DEP'T OF COMMERCE, INTERNET POL'Y TASK FORCE, COMMERCIAL DATA PRIVACY AND INNOVATION IN THE INTERNET ECONOMY: A DYNAMIC POLICY FRAMEWORK 34-35 (2010), http://www.commerce.gov/sites/default/files/documents/2010/december/iptf-privacy-green-paper.pdf.

[53] 45 CFR § 46.111.

[54] OFFICE OF MANAGEMENT & BUDGET, CIRCULAR NO. A-94 REVISED, MEMORANDUM FOR HEADS OF EXECUTIVE DEPARTMENTS AND ESTABLISHMENTS, GUIDELINES AND DISCOUNT RATES FOR BENEFIT-COST ANALYSIS OF FEDERAL PROGRAMS (Oct. 29, 1992), http://www.whitehouse.gov/omb/circulars_a094.

Conclusion

Big data provides a necessary catalyst for many important conversations about privacy. From advancing practical de-identification to infusing ethical decision-making into data projects, big data will require organizations to create new processes and innovative tools to engender trust and address privacy.

The Consumer Privacy Bill of Rights offers a helpful, flexible framework to resolve these challenges. More work is needed to ensure the principles in the Consumer Privacy Bill of Rights are operationalized by organizations in a common-sense fashion, but we already have a path forward to thoughtfully address the concerns posed by big data.

FPF thanks the National Telecommunications and Information Administration for considering these Comments, and we look forward to further engagement and collaboration on the issue of big data.

Sincerely,

Jules Polonetsky
Director and Co-Chair
Future of Privacy Forum

Christopher Wolf
Founder and Co-Chair
Future of Privacy Forum

Josh Harris
Policy Director
Future of Privacy Forum

Joseph Jerome
Policy Counsel
Future of Privacy Forum

Appendix A

Benefit-Risk Analysis for Big Data Projects

Jules Polonetsky
Omer Tene

FPF
FUTURE OF PRIVACY FORUM

Benefit-Risk Analysis for Big Data Projects

Jules Polonetsky
Omer Tene

August 2014

About the Future of Privacy Forum

Future of Privacy Forum (FPF) is a Washington, DC based think tank that seeks to advance responsible data practices. The forum is led by Internet privacy experts Jules Polonetsky and Christopher Wolf and includes an advisory board comprised of leading figures from industry, academica, law and advocacy groups.

To learn more about FPF, please visit www.futureofprivacy.org

Special thanks to The Privacy Projects for their support for this project.

Introduction: This analysis provides guidance for organizations in their weighing of the benefits of new or expanded data processing against attendant privacy risks.

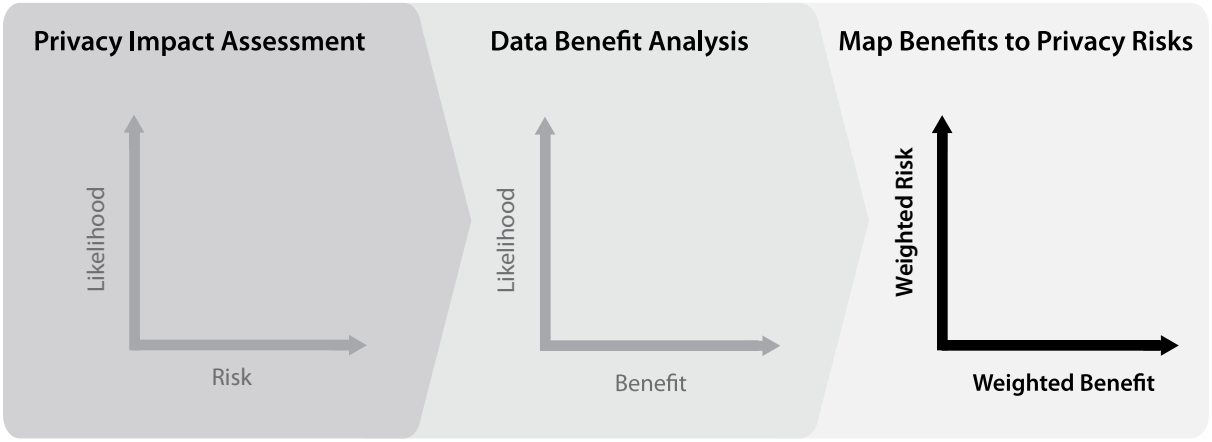
When responsible organizations identify new ways to process data, for example, when launching a new program, product, system or service, they utilize **Privacy Impact Assessments (PIA)** to conduct a systematic analysis to identify and address privacy issues. Current PIA practice includes detailed frameworks to help privacy professionals understand and quantify privacy risks.[1] Yet accounting for *risks* is only part of a balanced value equation. Decision-makers must also assess, prioritize, and to the extent possible, quantify a project's *benefits* in order to understand whether assuming the risk is ethical, fair, legitimate and cost-effective.

The phenomenon of "Big Data" exacerbates the tension between potential benefits and privacy risks by upping the ante on both sides of the equation. On the one hand, big data unleashes tremendous benefits not only to individuals but also to communities and society at large, including breakthroughs in health research, sustainable development, energy conservation and personalized marketing.[2] On the other hand, big data introduces new privacy and civil liberties concerns including high-tech profiling, automated decision-making, discrimination, and algorithmic inaccuracies or opacities that strain traditional legal protections.[3]

Decision-makers need to engage in a **Data Benefit Analysis (DBA).**

This document offers decision-makers a framework for a reasoned analysis to balance big data benefits against privacy risks. This process of identifying both benefits and risks is grounded in existing law. The Federal Trade Commission weighs benefits to consumers when evaluating the *unfairness* of business practices under Section 5 of the Federal Trade Commission Act. Similarly, the European Article 29 Data Protection Working Party applied a balancing test in its opinion interpreting the *legitimate interest* clause of the European Data Protection Directive.[4] The White House Office of Science and Technology Policy, which has recently studied the social and technical ramifications of big data, recognized the need to strike an appropriate balance between new opportunities and individual values.[5]

Structures and processes for sound benefit analysis are already well established. For example, in 1992, the White House Office of Management and Budget (OMB) issued guidelines for cost-benefit analysis of federal government programs and projects.[6] The OMB stressed that the criterion for deciding whether a government program can be justified is net present value, which is "computed by assigning monetary values to benefits and costs, discounting future benefits and costs using an appropriate discount rate, and subtracting the sum total of discounted costs from the sum total of discounted benefits." The OMB's guidance recognizes that some benefits may not be computable, but efforts to measure value can nevertheless produce useful insights. The same holds true with big data projects.



What is a Privacy Impact Assessment (PIA)?

A PIA is a decision-making tool used to identify and mitigate privacy risks at the beginning and throughout the development life cycle of a program, product, system or service.[7] While a formalized review process is not necessary for every use of data, particularly if the data is neither sensitive nor identifiable, a PIA process helps organizations understand what personal information they are collecting, how it will be used, stored, accessed and shared, and how privacy risks can be mitigated.[8]

PIA Goals

1) IDENTIFY privacy risks arising from the collection, storage, or dissemination of information in a potentially identifiable form.

2) EVALUATE compliance obligations and possible ways to mitigate privacy risks.

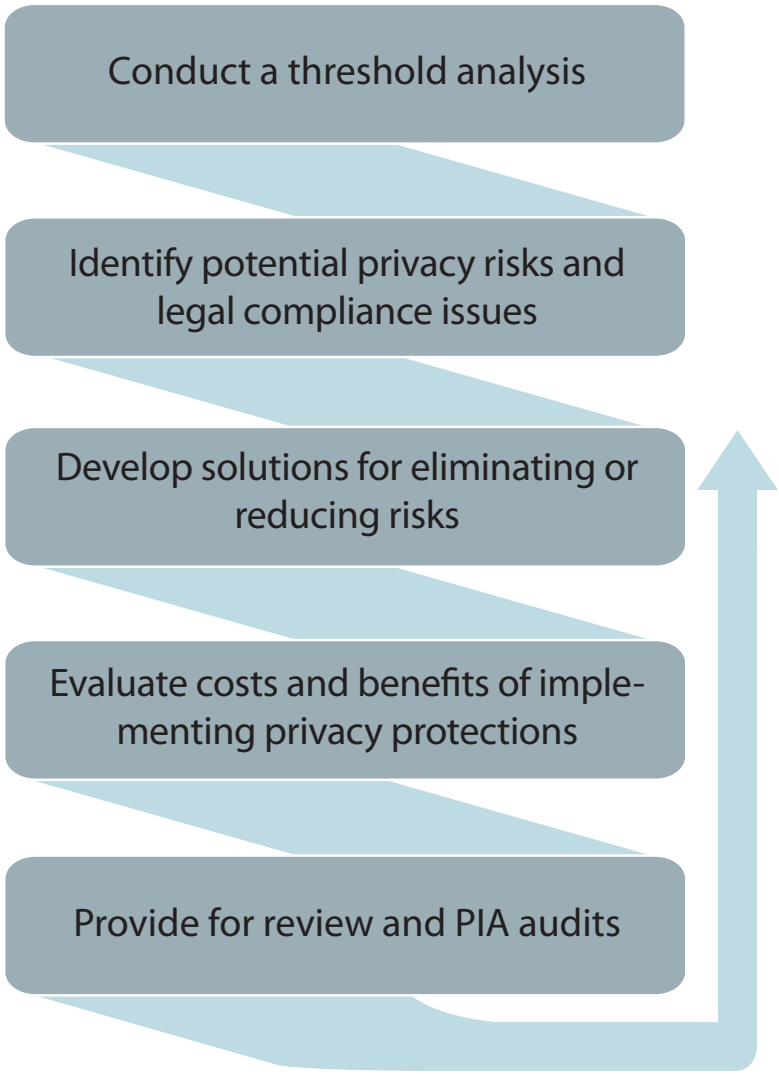
Many organizations have gained experience incorporating PIA into project management. A PIA is a necessary and proactive feature of managing risk in a responsible organization.

At its core, a PIA requires a value judgment to be made concerning the estimated **level of privacy risk** and the **likelihood** that such risk would materialize. In addition, a PIA involves determining how an organization can best employ risk mitigation tools to comply with privacy principles, generally captured in the **Fair Information Practice Principles (FIPPs),** as well as with individuals' expectations of privacy.

Mitigation measures can include policies around notice and choice, data minimization, or limited retention of data. A PIA allows an organization to calibrate its project to avoid using certain types of data or to further aggregate information being used in a project.

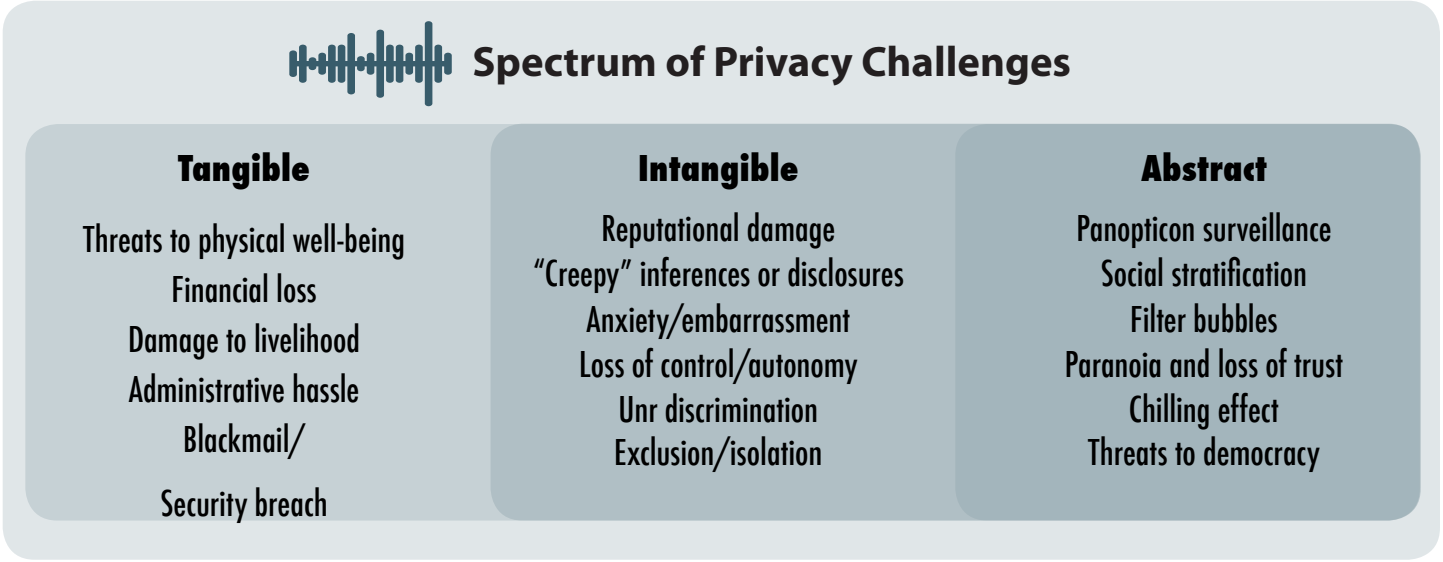
PIA Steps

A considerable body of knowledge has been created to help guide organizations through the implementation of a PIA. While specific PIA policies and procedures vary depending upon the nature of an organization and scope of data use, a typical PIA process consists of the following stages:[9]



Organizations have come to realize that privacy risk, sometimes conceptualized as *privacy harm*,[10] comes in different flavors. Various frameworks have developed to help organizations categorize privacy risk. Daniel Solove's taxonomy classifies privacy risks into four categories – information collection, information processing, information dissemination, and invasion – which, in turn, are broken out into 16 sub-categories.[11] Richard Thomas distinguishes between material privacy harms, moral privacy harms, and broader democratic and societal harms.

Big data presents new challenges impacting the entire risk spectrum. It accentuates not only the traditional tangible privacy harms but also the more abstract, ethical challenges requiring businesses and governments to make weighty value choices. Existing risk assessment frameworks are geared to identify and address tangible harms, such as financial loss or security vulnerabilities.



Yet big data situations require a broader view of risk as well as additional analysis of a project's ethical implications. As the risk taxonomy above suggests, some of the new risks commonly associated with big data are not easily mapped to any traditional recognizable harms. Other concerns are that data analysis could permit new forms of unfair discrimination, stigmatization and narrowcasting. All of these new concerns must also be incorporated into a PIA, which may therefore require careful consideration of a project's abstract or unintended consequences.

Many new uses of data conducted by organizations may be routine, involving simply a use that does not create new risks, or a use that is subject to well-defined measures that eliminate risk. The analysis documented here is necessary when a minimum threshold has been surpassed. It is critical for organizations to have personnel and processes in place that can spot new issues of concern and determine which issues should be subjected to a PIA and a benefit review.

Traditionally, organizations mitigated privacy risks by operationalizing the FIPPs, including enhancing notice and choice, limiting data retention, improving individuals' access and ensuring accountability. Alas, the onset of big data practices has introduced formidable challenges to some of these fundamental principles, including the scope of the framework (often addressed by defining the term "personally identifiable information" (PII)), the concepts of data minimization, purpose limitation and consent, and the right of individual access.[12] This has required policymakers and professionals to develop new privacy solutions to augment traditional tools. These enhanced solutions address the broader categories of privacy risks that are created by big data:[13]

Enhanced Transparency:

Like any interpretative process, big data analysis is prone to errors, inaccuracies and bias.[14] Consequently, organizations should provide more transparency into their automated processing operations and their decision-making processes, including eligibility factors and marketing profiles, in order to empower individuals.

Featurization:

Organizations should increase the ability of individuals to access to their data in intelligible, usable form in ways that allow them to analyze and utilize their own information. Featurization will allow individuals to declare their own policies, preferences and terms of engagement and "share the big data wealth" with organizations.

Privacy by Design:

Organizations should integrate privacy considerations early into lifecycle of new products and services. The assessment of privacy challenges at the design stage helps stem privacy risks at their outset, and privacy be design processes encourages organizations to revisit privacy issues throughout a project's life.

De-Identification:

While there are many different understandings of what constitutes effective anonymization, organizations should implement practical de-identification processes that make use of legal and administrative safeguards, in addition to reasonable technical measures. Both the sensitivity of the data and the utility of the data must be considered. Determining how to balance the utility of data against various threat models may itself require a benefit-risk analysis.

Risk mitigation strategies are essential for protecting privacy, yet at the same time they may constrain beneficial uses of data, thereby minimizing data utility. In addition, mitigation strategies alone do not help organizations decide **when is it worthwhile to proceed with a big data project despite residual privacy risks**. For example, if big data analysis can generate a health benefit that will improve the lives of millions of people, it may be ethical to allow a project to proceed even if privacy risks cannot be completely eliminated. Conversely, if the likelihood of accomplishing that benefit is extremely remote or if the contemplated benefit is minor, large privacy risks would not be justified.

By focusing exclusively on privacy risk, existing PIA practice does not account for the tremendous variance in anticipated big data benefits. This drives policymakers and corporate decision-makers into rote discussions of an almost ideological nature, with each side claiming the moral high ground and fully discounting arguments made by the other side. What is needed is a more thorough vocabulary of big data benefits. The following analysis proposes a methodology to better structure the discussion of big data benefits, assessing such variables as the ***nature*** of the benefit, the ***identity*** of the beneficiary and the ***likelihood*** of success. The results of this process, in turn, will feed into existing PIA practice to form a balanced, comprehensive view of big data risks and rewards.

Big data promises extraordinary benefits ranging from breakthroughs in medical research to enhancement of product offerings.

A Global Human Trafficking Hotline Network.[15] Non-profit organizations collaborate to establish an international information-sharing database to collect sensitive information about human trafficking. Together with law enforcement authorities, these organizations can use this information to help combat organized crime.

Internet Searches Reveal Harmful Drug Interactions. Medical researchers use massive datasets of de-identified Internet search results to discover harmful drug interactions, by comparing individuals' search queries against "fingerprints" of adverse side effects.[16]

Gathering Voice Data to Improve Speech Recognition.[17] Directory assistance services collect millions of voice samples in order to help create effective digital assistants that are embedded into mobile devices.

Ensuring High School Students Graduate.[18] Using data collected across school districts, analytic tools predict which students are at risk of failing or dropping out of school, providing schools and educators with a mechanism for early intervention.

Improve Newspapers Ability to Compete Online.[19] Traditional news publishers like the New York Times are turning to data in order to better serve subscribers with targeted reporting and interest-based content.

Location Data Create Intelligent Highways.[20] Geolocation information automatically generated by commuters' mobile devices is used to visualize traffic patterns in real time, helping urban planners manage traffic to deliver cost savings, environmental benefits, and a higher quality of life for commuters.

Using TV Viewing Habits to Create Better Entertainment.[21] By analyzing at the viewing habits of millions of users, entertainment streaming services can not only recommend better programming to viewers but also create new shows and programs that are better tailored to their viewers' tastes.

Tracking lost baggage and improving customer service.[22] Airlines are increasingly using data to offer flyers the ability to track their bags from curb to baggage claim, and big data analytics is directly improving customer experiences by allowing airlines to understand what travelers desire.

So far, there has been no procedural framework in place to assess big data benefits in a way commensurate with existing PIA risk frameworks. Yet accounting for *costs* is only part of a balanced and ethical value equation. In order to complete a cost-benefit analysis, organizations need to have at their disposal tools to assess, prioritize, and to the extent possible, quantify a project's *rewards*. Not all benefits are or should be treated as equal: a potentially big benefit with a high likelihood of success must be treated differently than a smaller benefit with a similarly high likelihood of success – or a big benefit that is unlikely to ever be accomplished.

The scope and dimensions of big data benefits have not been accounted for under current PIA practice. Yet existing legal frameworks already recognize the need to balance privacy risks against data rewards. For example, Section 5 of the Federal Trade Commission Act defines as "unfair" a practice that "causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves *and not outweighed by countervailing benefits to consumers or to competition*."[23] Similarly, the European Data Protection Directive[24] and new draft Regulation[25] authorize the processing of personal data based upon a "legitimate interest" of an organization, requiring organizations to perform a balancing test between individual risks and organizational rewards.

> The Federal Trade Commission engages in a balancing test when determining when a practice is "unfair" by assessing whether any potential injury to consumers is **not outweighed by countervailing benefits** to consumers.

> The Article 29 Working Party recognizes that in determining the "legitimate interests" of organizations, those interests must be weighed against **the potential effect on individual rights**.

The European Article 29 Data Protection Working Party has recently presented a balancing test to help organizations determine whether their legitimate interest in processing data outweighs the rights or interests of individual data subjects. The test recognizes that benefits may range "from insignificant through somewhat important to compelling.[26]

In similar vein, institutional review boards, which evaluate the ethics of human subject research proposals, are instructed to evaluate risks in relation to anticipated benefits, taking into account both prevailing community standards and subjective risk and benefit determinations.[27]

Maximizing the potential of big data requires a new framework for evaluating not only the risks but also the benefits created by novel data uses. To account for unique big data benefits and risks, organizations should engage in the following data benefit analysis.

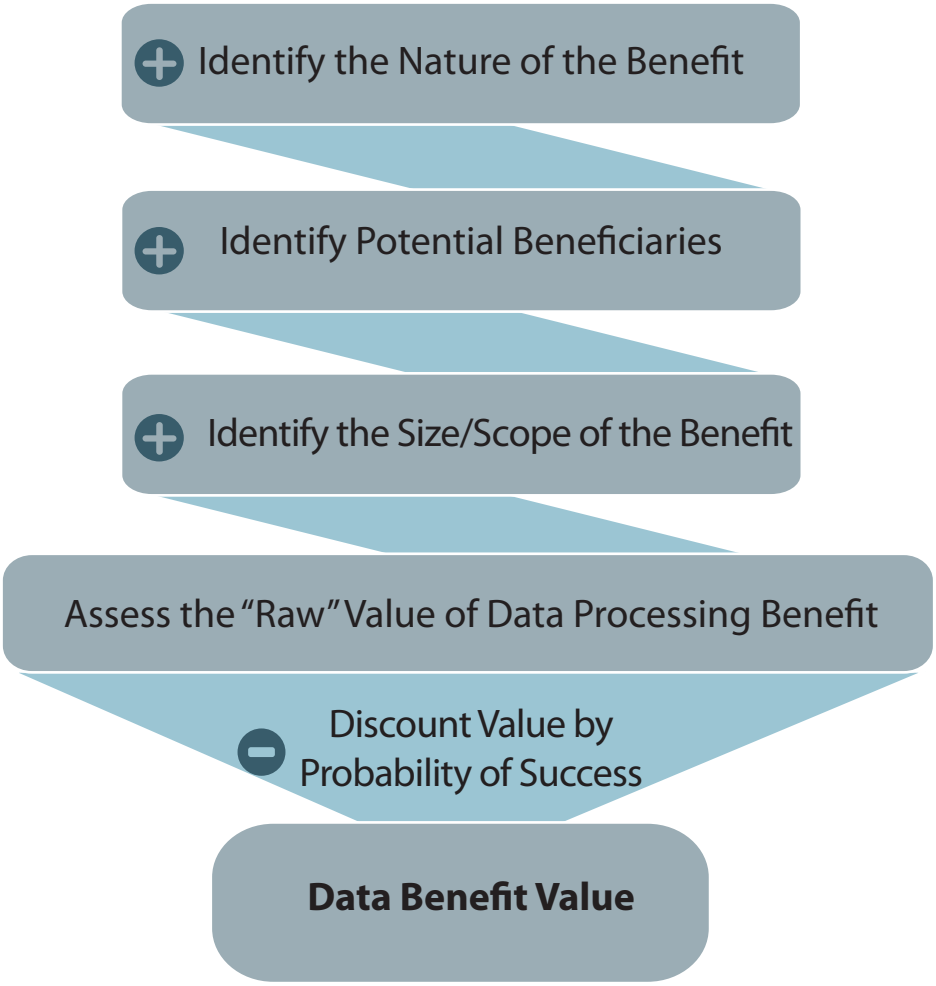
Data Benefit Analysis comprises two elements. **First**, organizations should assess the "raw value" of the benefit, which consists of (1) the nature of the benefit, (2) the potential beneficiaries, and (3) the degree (or size and scope) of the benefit. **Second**, organizations should discount the raw value score by the probability that it can achieved to obtain a discounted value score.[28]

This process intertwines with and complements any risk assessments. The end goal is to achieve an optimal balance between organizational needs and individual privacy.

There are no definitive rules on what degree or probability of benefit is needed to overcome presumptions against creating privacy risk. It is clear that the mere assertion that a product or service can be improved is not enough; yet proof beyond any doubt is an impossible standard.

Any analysis must take into account culture-specific differences in evaluating the relative weight of each parameter. For example, the relative value of a health or national security benefit may differ from society to society. Some societies may place a high value on individual benefits, while others may give greater weight to community values.

Depending upon how each of these factors compute, an organization will compile a raw value that reflects the potential benefit of a project – before taking into account uncertainty and weighing the benefits againstprivacy risks.



Hypothetical Case Study

Acme Corporation develops Road Runner, a fitness app that collects and analyzes information about users' diet, health, exercise and sleep. The app's data analysis provides users with helpful insights about their lifestyle, enabling them to optimize calorie consumption, reduce blood sugar and cholesterol levels, create a balanced exercise schedule, comply with doctors' prescriptions, and more.

A free app, Road Runner quickly gains traction, achieving a strong following with millions of users across the world. Acme collects and stores granular information about Road Runner users' habits, compiling statistics and creating graphs and indices that are accessible by users through an easy to use dashboard. In addition, Road Runner gives users real time notifications to inform them of any developing health conditions, such as lack of sleep, hypertension, or dehydration, or failure to take medication.

Acme incentivizes employers to pre-package Road Runner into their mobile application management platforms by promising potential savings on their health care benefit plans. In turn, some employers are offering bonuses to employees who lose weight, optimize their body fat percentage, or exercise more. In addition, healthcare providers are urged to recommend to their patients usage of the app to enhance adherence with prescription medicine regimens.

Acme retains user data indefinitely, but keeps it in de-identified form by assigning random identifiers to individual users. Acme's CEO argues that in the future, the data retained by the company could be used to prevent epidemics, cure lethal disease, and increase life expectancy by up to 40 years.

To help the research community, Acme provides health researchers in accredited schools with access to its information. According to a recent article in the American Journal of Medication, researchers have been able to utilize Road Runner data to find a concealed harmful interaction between two best-selling drugs.

In the U.S., Acme provides periodic reports on longitudinal studies about users' health and behavior to the U.S. Department of Health and Human Services. The reports, which are in aggregated form, help the federal government make decisions concerning public health and research funding.

While Road Runner users' de-identified information could conceivably be linked to PII with varying degrees of certainty, this would have to be done through highly complex (and expensive) processes of data matching and analysis, which neither Acme nor its researchers and business partners have a clear interest to partake.

How to approach:

1. Conduct a full assessment of the **benefits** of a proposed data project (see pages 7-10)
2. Recognize and account for traditional **privacy risks** as well as risks that are unique to big data, and explore strategies for **risk mitigation** (see pages 2-4)
3. **Weigh** unmitigated privacy risks against big data rewards and determine how to proceed with a project (see page 11)

1) Identify the nature of the benefit:

Big data projects increasingly promise wide-ranging benefits to scientific research, public health, national security, law enforcement, energy conservation and economic efficiency. Organizations should recognize that the nature of the benefit must be accounted for by an analysis that measures social and cultural priorities and sensitivities.

The Road Runner app promises better information about personal health, cost savings to companies and communities, and additional knowledge to help inform public health policy and funding decisions.

2) Identify the potential beneficiaries:

Data projects can affect a wide-variety of stakeholders. These include not only the individual whose data is processed and the business that is processing the information, but increasingly also the government, a community, or society at large. As the OMB explains, "Analyses should include comprehensive estimates of the expected benefits and costs to society based on established definitions and practices for program and policy evaluation."[29]

The Road Runner app primarily benefits the individual whose data it collects; but Acme also promises benefits to organizations through insurance cost savings, as well as to government, the research community, and potentially, society at large..

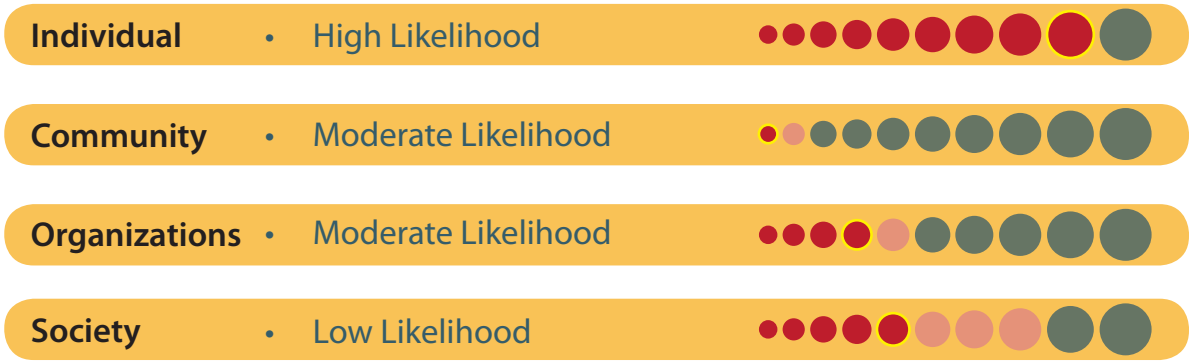
3) Assess the size or scope of the benefit and assign a Raw Value Score:

Individual	•	Better information about personal health	
Community	•	Healthier, more active individuals	
Organizations	•	Savings on health care benefit plans	
Society	•	More informed research and funding prioritization	

A raw value score combines the assessment of the beneficiary the nature, size and scope of the benefit. The raw value represents the absolute value of a project prior to its discounting by probability and risk.

4) Discount by the probability of success:

After computing a raw value score, an organization must assess the likelihood that the benefits of a project will in fact come to pass. Uncertainty constitutes a **discount factor** that reduces the initial raw value score. **The certainty of obtaining the desired benefit is an essential element in determining the desirability of assuming related privacy risks.**



The Data Benefit Analysis should not be viewed as a static framing exercise. In many cases, mitigation techniques may impact data utility by reducing the potential benefit. This means that the **Data Benefit Analysis** is a dynamic process, through which mitigation techniques are carefully calibrated to optimize the risk-benefit equation in order to reach the apex point. The OMB calls this exercise "sensitivity analysis," noting that "[m]ajor assumptions should be varied and net present value and other outcomes recomputed to determine how sensitive outcomes are to changes in the assumptions."[30] Of course, in many cases, a baseline level of protection against risk will mandatory under regulation in order to support the legitimacy of the data processing.

If data are further de-identified, societal benefit is reduced but so is attendant privacy risk.

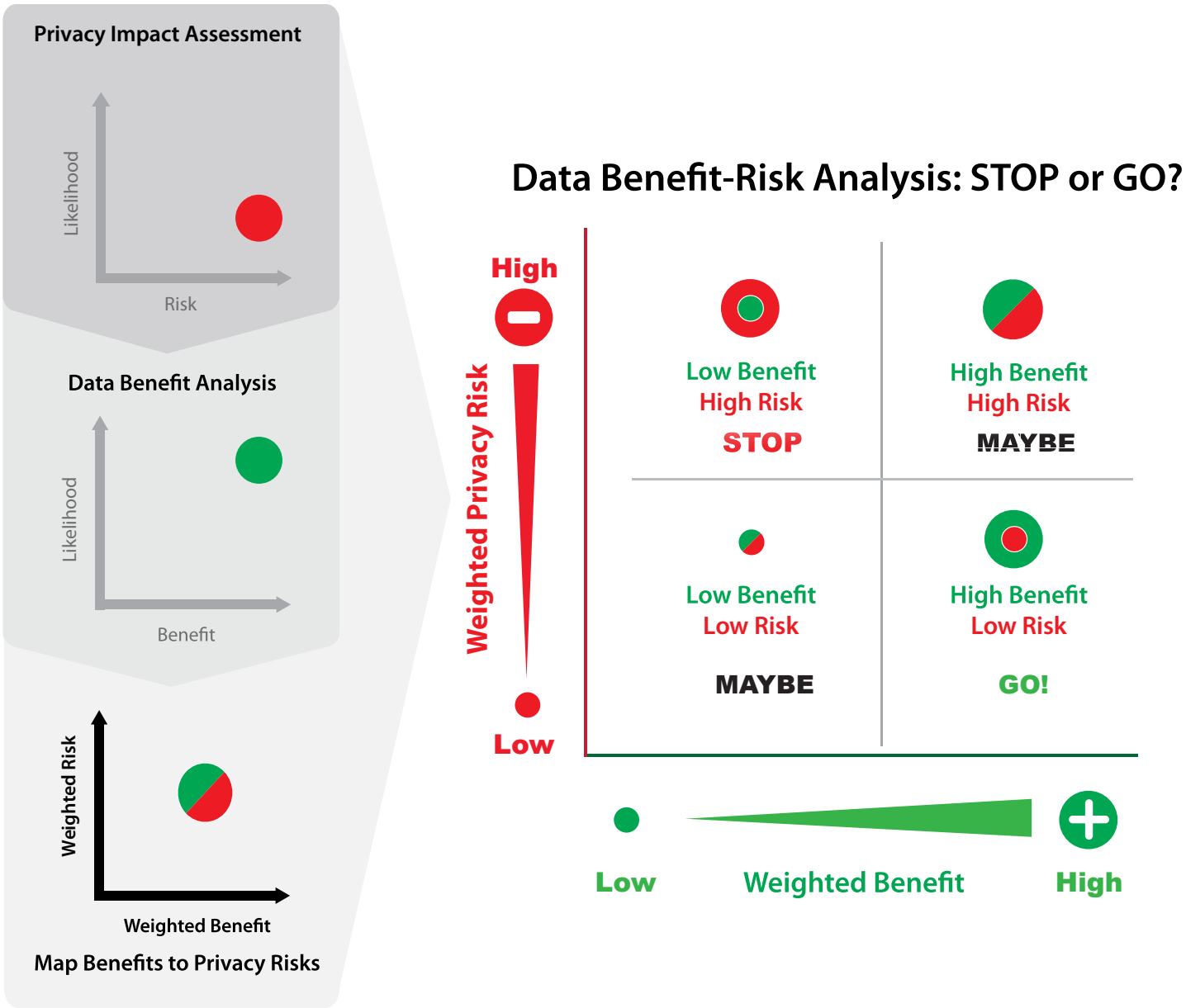
Personalized learning algorithms, for example, can be validated to minimize the risk of inaccuracy, but this may lower societal benefit by reducing the likelihood of serendipitous findings.

Mapping Benefits against Risks

Once an organization has a better understanding of a project's benefits, it can map the **discounted benefit value** against privacy risks identified through a PIA. By doing so, it can now visualize a complete picture to inform decision-making weighing both benefits and risks.

By mapping benefits against risks, an organization evaluates the merits of a big data project. To do so, an organization must elucidate where a project falls on the a risk-benefit continuum.

Mapped in this way, a contemplated project is placed on a continuum ranging from projects that the FTC and the Article 29 Working Party may view as unfair to project that the regulators view as being within the legitimate interest



While some of the assessments proposed in this framework can be standardized and quantified, others require value judgments and input from experts other than privacy professionals or data regulators. For example, assessing the scientific likelihood of capturing a benefit in a specialized area cannot be made solely based on privacy expertise. This framework cannot achieve mathematical accuracy given the inherent degree of subjectivity in assessing the relative erits of various benefits. However, this has not stopped policymakers in other arenas from proposing structured processes to measure project benefits against risks. For example, the OMB states, "Although net present value is not always computable … efforts to measure it can produce useful insights even when the monetary values of some benefits or costs cannot be determined."[31]

This highlights the importance of determining *who* will be tasked with undertaking the cost-benefit analysis. Moving forward, organizations will need to create or expand accountable data ethics review processes to engender trust and address privacy. Many companies have already laid the groundwork to address these decision-making challenges by appointing Chief Privacy Officers or building internal ethical review programs. Further efforts are needed to understand the most effective structures for different organizations and different types of data. Models may range from a formal Institutional Review Board-type process to empowering Chief Privacy Officers through cross-functioning privacy committees, or involve building structures such as extra advisory boards or opportunities for policy maker or regulator input.

What is an Institutional Review Board?[32]

Institutional Review Boards (IRB) emerged as the chief regulatory response to concerns about ethical abuse in the use of human subjects for research. IRBs are therefore charged with balancing the potential risks and benefits arising from any project involving human subject research. Policy guidance on IRBs recognizes that research benefits fall into different categories, including acquiring new knowledge, improving drug safety, promoting technological advances, or providing better healthcare.

IRBs must have at least five members, encompassing a wide-variety of backgrounds and professional expertise. Boards that review research involving specific categories of human subjects, such as children, pregnant women, or the mentally disabled, must include members who have special experience with those groups.

An IRB's final assessment of a project depends on prevailing community standards and subjective determinations of risks and benefits. While there are limits on the risks that individuals should ethically be asked to accept for the potential benefit of others, IRBs are generally directed to not be overprotective.

End Notes

1. Trilateral Research & Consulting, Privacy impact assessment and risk management, A Report for the Information Commissioner's Office (May 4, 2013), http://ico.org.uk/about_us/consultations/~/media/documents/library/Corporate/Research_and_reports/pia-and-risk-management-full-report-for-the-ico.pdf.
2. Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 Stan. L. Rev. Online 63 (2012).
3. Civil Rights Principles for the Era of Big Data, http://www.civilrights.org/press/2014/civil-rights-principles-big-data.html (last visited March 15, 2014).
4. Article 29 Data Protection Working Party, Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC (Apr. 9, 2014), https://www.huntonprivacyblog.com/files/2014/04/wp217_en.pdf.
5. White House, Executive Office of the President, Big Data: Seizing Opportunities, Preserving Values (May 2014), http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.
6. Office of Management & Budget, Circular No. A-94 Revised, Memorandum for Heads of Executive Departments and Establishments, Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs (Oct. 29, 1992), http://www.whitehouse.gov/omb/circulars_a094.
7. *See generally*, Information Commissioner's Office, Conducting Privacy Impact Assessments Code Of Practice (February 2014), http://ico.org.uk/for_organisations/data_protection/topic_guides/~/media/documents/library/Data_Protection/Practical_application/pia-code-of-practice-final-draft.pdf.
8. Department of Homeland Security, Privacy Impact Assessments, The Privacy Office Official Guidance (June 2010), https://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_guidance_june2010.pdf.
9. *See, e.g.*, ICO Privacy Impact Assessment Handbook (2010) (describing a five-part PIA process including (1) a preliminary phase, (2) preparation, (3) consultation and analysis, (4) documentation, and (5) review and audit. *See also* http://www.piafproject.eu (last visited March 15, 2014).
10. Ryan Calo, *The Boundaries of Privacy Harm*, 86 Ind. L.J. 1131 (2011); Silicon Flatirons Center, The New Frontiers of Privacy Harm, January 2014, http://www.siliconflatirons.org/events.php?id=1381.
11. Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. Pa. L. Rev. 477 (2006). *Information collection* includes surveillance and interrogation; *information processing* includes aggregation, identification, insecurity, secondary use and exclusion; *information dissemination* includes breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation and distortion; and *invasion* includes intrusion and decisional interference.
12. Julie Brill, Commissioner, Fed. Trade Comm'n, Remarks at Fordham University School of Law: Big Data, Big Issues (Mar. 2, 2012) (transcript available at http://ftc.gov/speeches/brill/120228fordhamlawschool.pdf). Commissioner Brill said: "Big Data's impact on privacy is requiring some new and hard thinking by all of us."
13. Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 Nw. J. Tech. & Intell. Prop. 239 (2013).
14. *Id.* at 271.
15. Press Release, Polaris Project Launches Global Human Trafficking Hotline Network, Polaris Project (Apr. 9, 2013), http://www.polarisproject.org/media-center/news-and-press/press-releases/767-polaris-project-launches-global-human-trafficking-hotline-network-.
16. Nicholas Tatonetti et al., *Detecting Drug Interactions From Adverse-Event Reports: Interaction Between Paroxetine and Pravastatin Increases Blood Glucose Levels*, 90 Clinical Pharmacology & Therapeutics 133 (2011); Nicholas Tatonetti et al., *A Novel Signal Detection Algorithm for Identifying Hidden Drug-Drug Interactions in Adverse Event Reports*, 12 J. Am. Med. Informatics Ass'n 79 (2011).
17. Derrick Harris, *Google Explains How More Data Means Better Speech Recognition*, GigaOM (Oct. 31, 2012), http://gigaom.com/2012/10/31/google-explains-how-more-data-means-better-speech-recognition/.
18. Press Release, Alabama's Largest School District Turns to IBM to Build a Smarter Education System, IBM (July 13, 2009), http://www-03.ibm.com/press/us/en/pressrelease/27984.wss.
19. Antonio Regalado, *Unsubscribing? The New York Times Wants to Predict That*, MIT Tech. Rev. (Feb. 12, 2014), http://www.technologyreview.com/news/524716/unsubscribing-the-new-york-times-wants-to-predict-that/.
20. Haomiao Huang, *Calling All Cars: Cell Phone Networks and the Future of Traffic*, ArsTechnica (Feb. 24, 2011), http://arstechnica.com/gadgets/2011/02/calling-all-cars-measuring-traffic-using-cell-phone-data/.
21. David Carr, *Giving Viewers What They Want*, N.Y. Times (Feb. 24, 2013), http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html?pagewanted=all&_r=1&.
22. Katherine Noyes, *For the Airline Industry, Big Data Is Cleared for Take-Off*, Fortune (June 19, 2014), http://fortune.com/2014/06/19/big-data-airline-industry/.
23. 15 U.S.C. Sec. 45(n) (emphasis added).
24. Article 7(f) of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995 O.J. (L 281) 31 (Nov. 23, 1995), available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1995:281:0031:0050:EN:PDF.
25. Article 6(1)(f) of Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final (Jan. 25, 2012), available at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf.
26. Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller Under Article 7 of Directive 95/46/EC, 844/14/EN WP 217, at 30 (Apr. 2014).
27. In its guidance, the OMB calls for discounting both costs and benefits, stating that "Estimates of benefits and costs are typically uncertain because of imprecision in both underlying data and modeling assumptions. Because such uncertainty is basic to many analyses, its effects should be analyzed and reported."
28. Office of Management & Budget, Circular No. A-94 Revised, *supra* note 6.
29. *Id.* at 6, *supra* note 6.
30. *Id.* at 9(c), *supra* note 6.
31. *Id.* at 5(a), *supra* note 6.
32. 45 CFR § 46. *See also* U.S. Dep't of Health & Human Resources, Institutional Review Board Guidebook (1993), *available at* http://www.hhs.gov/ohrp/archive/irb/irb_chapter3.htm#e1.


FPF
FUTURE OF
PRIVACY FORUM