FUTURE OF PRIVACY FORUM

October 15, 2014

via email: privacyeng@nist.gov

Ms. Diane Honeycutt
Secretary
Computer Security Division
National Institute of Standards and Technology
100 Bureau Drive, Stop 8930
Gaithersburg, MD 20899

Re: Public Comments, NIST Privacy Engineering Workshop

Dear Ms. Honeycutt:

The Future of Privacy Forum (FPF) is a think tank seeking to advance responsible data practices and is supported by leaders in business, academia and consumer advocacy.[1] FPF thanks the National Institute of Standards and Technology (NIST) for providing this opportunity to comment upon the NIST Privacy Engineering Objectives and Risk Model Discussion Draft.

NIST has long played an important role in developing technical research and standards in the United States. We thank NIST for providing a forum for diverse stakeholders to identify and discuss critical issues in cybsersecurity and privacy engineering. In its first two workshops on privacy engineering, NIST has begun the difficult but fundamental work of bridging communication gaps that hinder privacy operationalization. FPF shares NIST's desire to instill the complex world of privacy engineering with more consistent objectives and terminology. However, we propose what we believe will be a more effective way for NIST to advance a structure of privacy engineering objectives, as an alternative to the Discussion Draft.

First, the current Discussion Draft and Discussion Deck propose design objectives and a risk model based primarily on NIST staff's own review of "current and long-standing theories on the concept of privacy."[2] While NIST's search for overarching system characteristics that encompass a variety of privacy concepts is sound, we believe that its current draft incorporates too much of unsettled public policy debates. In the absence of clear consensus or guidance from policymakers, the explicit inclusion of unclear privacy harms, engineering objectives and risk terminology into a NIST Interagency Report (NISTIR) will only muddy the waters and complicate compliance efforts.

Rather than reinvent privacy conceptions from the ground up, we urge NIST to look to more settled privacy frameworks for guidance. Existing frameworks, such as the White House's Framework for

[1] The views herein do not necessarily reflect those of the Advisory Board or supporters of the Future of Privacy Forum.
[2] NIST, NIST Privacy Engineering Objectives and Risk Model Discussion Draft, *available at* http://www.nist.gov/itl/csd/upload/nist_privacy_engr_objectives_risk_model_discussion_draft.pdf.

Protecting Privacy and Promoting Innovation in the Global Digital Economy,[3] build upon the well-established Fair Information Privacy Principles (FIPPs). These standards are already widely understood and reflect generally accepted legal and policy objectives. Moreover, while FIPPs-based frameworks may be strained by our increasingly networked society, privacy engineering efforts could provide effective operational solutions while policy consensus develops. As we have argued before,[4] flexibility in the interpretation and application of the FIPPs is necessary to achieve a balance between consumer protection and today's rapid innovation. Nevertheless, the FIPPs themselves are and should remain at the heart of privacy programming.

If NIST were to adjust its objectives and risk model to correlate more directly with the FIPPs, then its privacy engineering efforts could be more productively directed towards designing practical information systems and controls capable of supporting existing privacy standards. FIPPs-based frameworks have been embraced by companies and countries around the world, not only supporting local privacy by design efforts but enabling greater global interoperability.

To illustrate, an information system whose objective is to achieve "Transparency"[5] would need to be capable of providing notices to consumers about an organization's privacy and security practices. Depending on the organization's own privacy standards and commitments, those notices might need to appear in different formats, or at different times. The system could then be engineered to, *e.g.*, associate certain data sets with certain notices, to deliver the notices and to confirm their delivery. In this way, the technical systems and standards would enable the organization to engage in privacy by design, protecting consumer privacy while making purposeful decisions about resource allocation and implementing effective, measurable and consistent controls.[6] (An illustrative map of FIPPs to potential processes follows in Appendix A).

Importantly, this mapping of accepted privacy standards to technical controls would also provide organizations with the flexibility to adapt to the new privacy risks and opportunities emerging from Big Data and the Internet of Things. For example, the White House Consumer Bill of Rights recognizes "Respect for Context" as a core privacy objective, based on the traditional FIPPs of purpose specification and use limitation.[7] While the information system should not dictate what "context" means for a particular consumer interaction, it could give organizations the capability to categorize and link data to permissible uses given a range of factors.

A system with these capabilities would support flexible application of FIPPs standards and also facilitate benefit-risk analyses, both of which FPF believes are critical to unlocking the benefits of Big Data.[8] For example, a system could be engineered to allow companies to systematically connect *out*-of-context uses

---

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

[4] *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.

[5] As defined by the White House Consumer Bill of Rights, for example. *See* WHITE HOUSE, *supra* note 3, at 14.

[6] *See* NIST, *supra* note 2.

[7] *See* WHITE HOUSE, *supra* note 3, at 50.

[8] *See, e.g.*, Comments of the Future of Privacy Forum RE: Big Data: A Tool for Inclusion or Exclusion? Workshop, Project No. P145406 (Aug. 2014), *available at* http://www.ftc.gov/system/files/documents/public_comments/2014/08/00027-92420.pdf.

to additional privacy protections, if the benefits of using the data in a novel way were demonstrated to outweigh the privacy risks. (A more in-depth description of our proposed Benefit Risk Analysis model is attached as Appendix B).

Next, in addition to shifting its efforts to *developing* technical processes that support organizations' existing privacy standards and commitments, we urge NIST to *highlight* technical processes and controls that already support such privacy objectives. In some more mature or highly regulated industries, exemplars of privacy by design standards and systems may already be widely available. NIST's ability to call attention to arenas where privacy-protecting technologies and standards have already been implemented would immediately benefit less experienced organizations and industries. Existing information tools, such as access controls, profile managers and retention and deletion schedules are just a few examples of systems that support privacy objectives that can be referenced here. By the same token, NIST could also shed light on practice areas where effective and easily accessible standards have *not* yet appeared. By more closely examining the existing landscape of privacy-supporting systems and technologies, NIST could identify areas where additional research may be necessary and where new technical processes would be most effective.

Finally, we recognize that NIST's technical expertise may prove most valuable in support of specific privacy-protecting controls, and would recommend that NIST consider engaging more deeply with de-identification. Although de-identification processes are powerful privacy controls, many organizations have found them difficult to operationalize. Continued uncertainty about the sufficiency of various technical and administrative de-identification techniques have left many organizations reluctant to invest in it. If NIST were to focus on identifying the range of industry practices used for various de-identification models and help inform efforts to design practical de-identification standards, more widespread adoption of these valuable techniques would follow.

In conclusion, we recommend that NIST shift the focus of its privacy engineering efforts prior to the development of the proposed NISTIR. Rather than creating a new privacy framework, NIST should encourage the development of technical processes capable of supporting established, flexible privacy standards and policies. Companies today have already committed themselves to upholding a range of privacy standards, which may arise from statutory obligations, self-regulatory regimes or binding public promises. What they need now are the technical tools and systems to ensure that they can live up those commitments.

We thank NIST for considering these comments.

Respectfully submitted,

Jules Polonetsky
Co-Chair and Director
Future of Privacy Forum

Christopher Wolf
Founder and Co-Chair
Future of Privacy Forum

Kelsey Finch
Policy Counsel
Future of Privacy Forum

Appendix A:

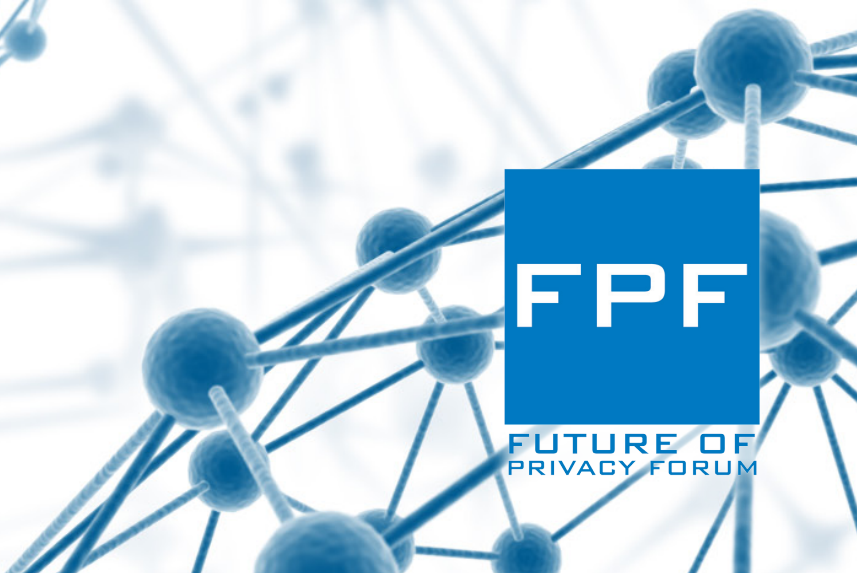
For illustrative purposes only, this table maps potential information system controls and capabilities against an existing privacy framework. This example utilizes the White House Consumer Bill of Rights' interpretation of the FIPPs for its privacy objectives.

Consumer Bill of Rights	Possible Controls and Capabilities
Individual Control	• System supports a range of choice mechanisms (e.g., more granular controls for more sensitive data, broader controls for less linkable data) • System tags data sets subject to specific choice mechanisms • System updates data tags automatically if consumer withdraws or limits consent (where the ability to withdraw or limit consent is necessary) • If necessary, system provides mechanisms for withdrawing or limiting consent
Transparency	• System supports a range of notice mechanisms (e.g., short-form notices, pop-up warnings) • System tags data sets subject to specific notice requirements System verifies and records that notice was provided • System is structured so as to enable visibility into data use practices.
Respect for Context	• System manages data sets according to ranges of permissible purposes • System flags or restricts impermissible uses • System verifies and records approved uses
Security	• System supports a range of security safeguards (e.g., access to consumer financial information restricted, consumer cookie identifiers hashable) • System tags data sets subject to particular privacy or security risks (e.g., sensitive data destroyed after certain time period)
Access and Accuracy	• System is capable of providing access to data sets when required • System permits data records to be corrected, if necessary • System removes inaccurate or outdated data • System permits data records to be provided in usable format to users, if necessary • System enables users access to their records, if necessary
Focused Collection	• System supports data retention and deletion schedules • System supports de-identification, if necessary • System tags and protects information organization has a legal obligation to retain • System tags data according to ranges of permissible purposes • System does not collect unintended data
Accountability	• System supports audit capabilities • System is capable of recording and reporting certain activities (e.g., modifications to data records are reviewable)

Appendix B:

Benefit-Risk Analysis for Big Data Projects

Jules Polonetsky
Omer Tene
Joseph Jerome



FPF
FUTURE OF PRIVACY FORUM

Benefit-Risk Analysis for Big Data Projects

Jules Polonetsky
Omer Tene
Joseph Jerome

September 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.

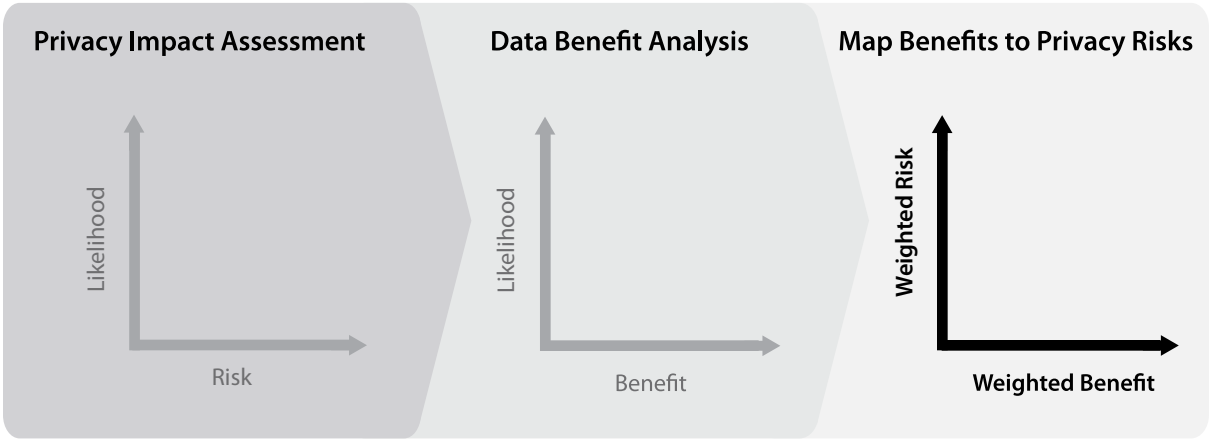
When responsible organizations identify new ways to process data, for example, when launching a new program, product, system or service, they utilize **Privacy Impact Assessments (PIA)** to conduct a systematic analysis to identify and address privacy issues. Current PIA practice includes detailed frameworks 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.

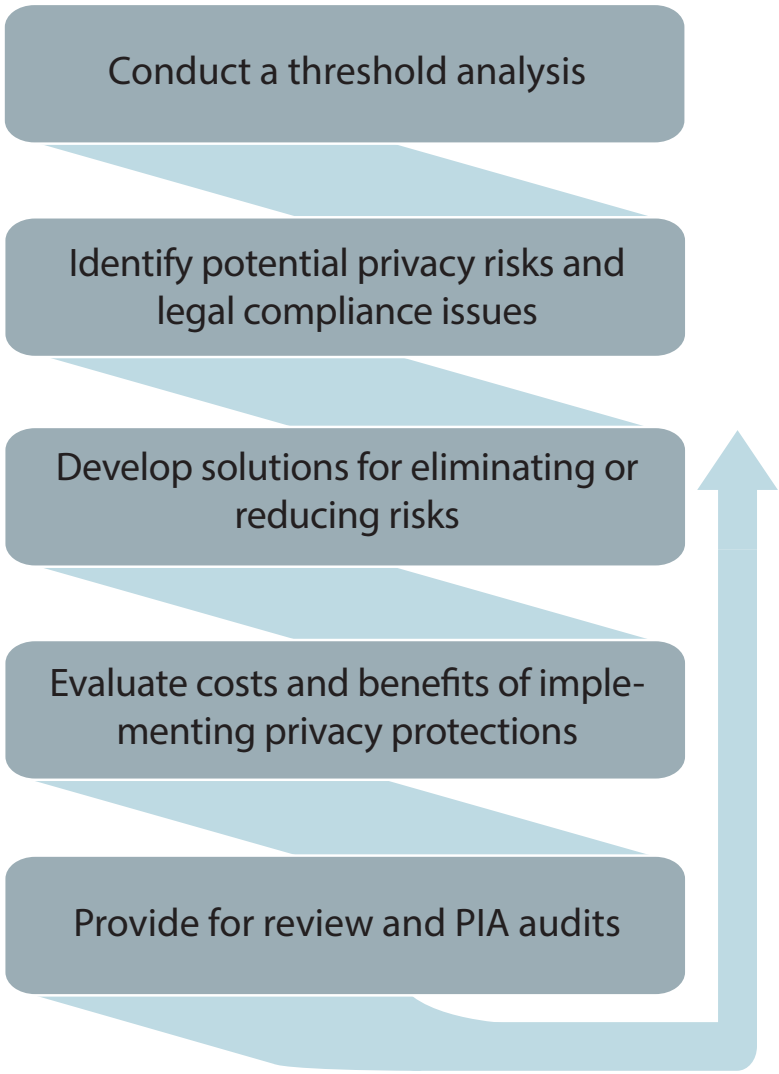
Many organizations have gained experience incorporating PIA into project management. A PIA is a necessary and proactive feature of managing risk in a responsible organization.

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 adjust 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.

Spectrum of Privacy Challenges

Tangible	Intangible	Abstract
Threats to physical well-being	Reputational damage	Panoptic surveillance
Financial loss	"Creepy" inferences	Social stratification
Damage to livelihood	Anxiety/embarrassment	Filter bubbles
Administrative inconvenience	Loss of control or autonomy	Paranoia and loss of trust
Security breach	Unfair discrimination	Chilling effect
Confidentiality breach	Exclusion and isolation	Threats to democracy

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 continue to be routine, involving uses of data that do not create new risks, or uses that are subject to well-defined measures for risk elimination. The analysis documented here is required when a minimum threshold is surpassed. Organizations must have personnel and processes in place to spot new issues and concerns and select those issues that should be subjected to further **PIA** and **Data Benefit Analysis**.

Risk Mitigation Strategies

Traditionally, organizations mitigated privacy risks by operationalizing the Fair Information Practice Principles (FIPPS), including enhancing notice and choice or limiting data retention. 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 decision-making processes, including eligibility factors and marketing profiles.

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. Moreover, privacy by design processes encourage 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.

Risk mitigation strategies are essential for protecting privacy, yet at the same time they can 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 a 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 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 tailored to their viewers' tastes.

Tracking lost baggage and improving customer service.[22] Airlines are increasingly using data to offer travelers 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 want.

So far, there has been no 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 tools to help them 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.

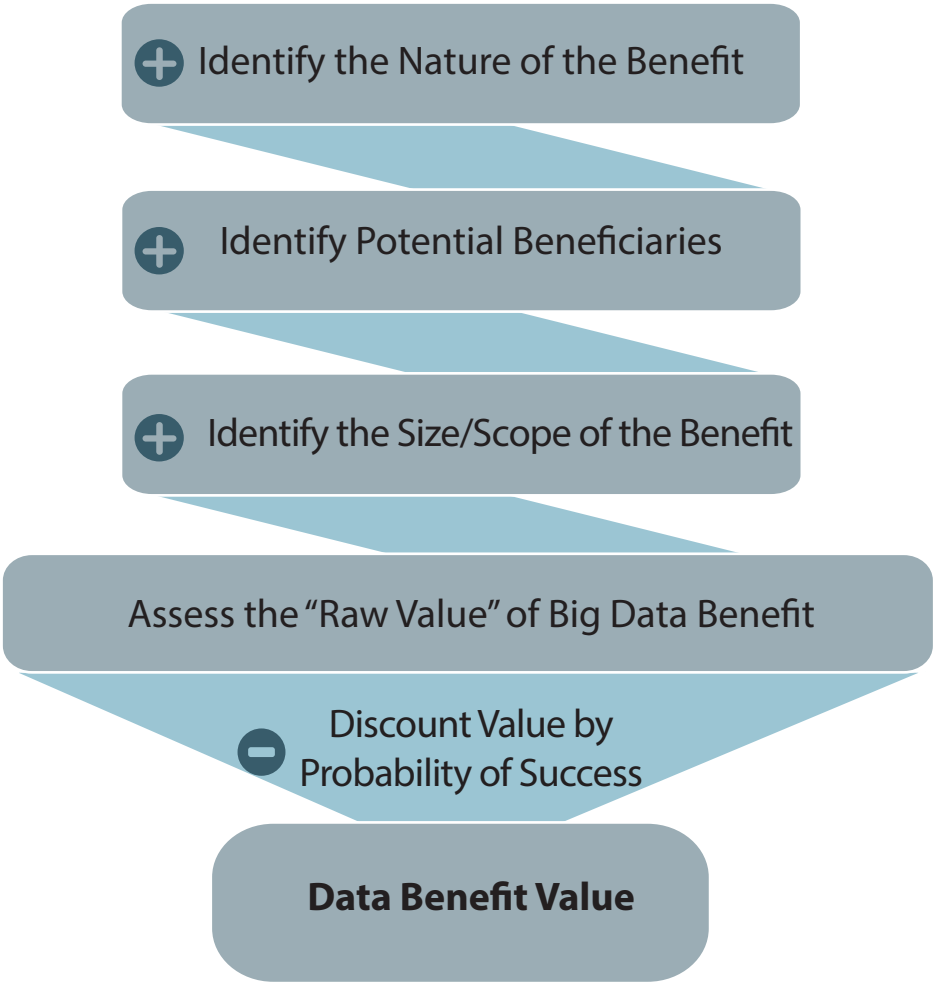
Data Benefit Analysis comprises two elements. **First**, organizations should assess the "raw value" of a benefit, which consists of (1) the nature of the benefit, (2) the potential beneficiaries, and (3) the degree (or size and scope) of the benefit. **Second**, organizations should discount the raw value score by the probability that the benefit can be 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 unreasonable 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 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 against privacy 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, 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 to 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 the proposed data project (see pages 9-10)
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 to 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 score represents the absolute value of a project prior to its discounting by probability and accounting for 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.**

Individual	•	High Likelihood
Community	•	Moderate Likelihood
Organizations	•	Moderate Likelihood
Society	•	Low Likelihood

The Data Benefit Analysis should not be viewed as a static framing exercise.

The ultimate value of a data project's benefits as well as the magnitude of its privacy asks are linked to the risk mitigation strategies that have been implemented.

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.

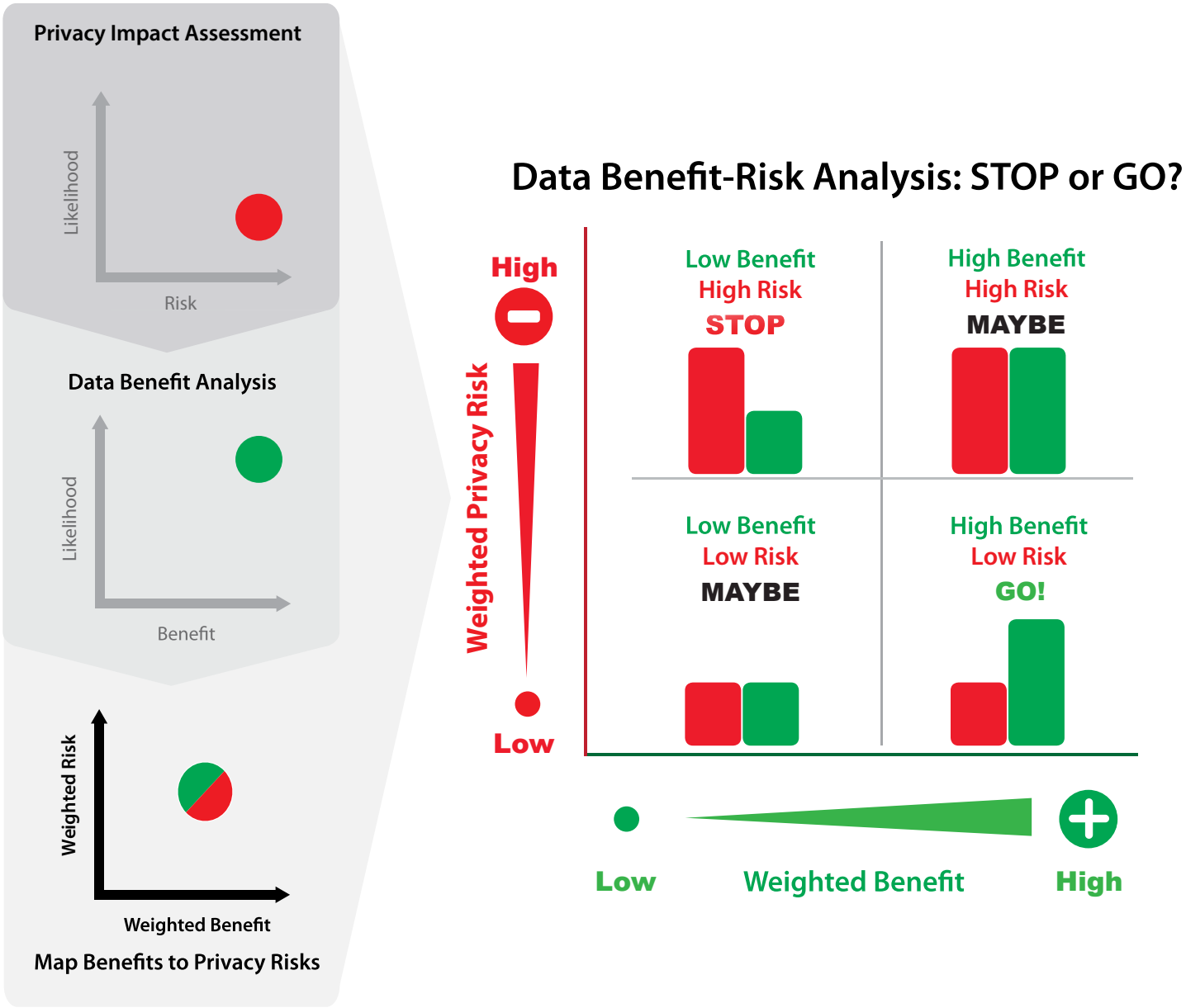
For example, as data are further de-identified, societal benefit is reduced but so is attendant privacy risk.

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 regard as being within the legitimate interest of the organization or the public at large.



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.

Furthermore, this framework cannot achieve mathematical accuracy given the inherent degree of subjectivity in assessing the relative merits 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 Data Benefit Analysis. Moving forward, organizations will need to create or expand accountable data ethics review processes to engender trust and address privacy concerns. 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 external advisory boards or opportunities for policy maker or regulator input.

What is an Institutional Review Board?

Institutional Review Boards (IRB)[32] 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. 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. 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