Software & Information
Industry Association
1090 Vermont Ave NW Sixth Floor
Washington, DC 20005-4095

SIIA

Submitted via email: privacyrfc@ntia.doc.gov

August 5, 2014

John B. Morris
Associate Administrator and Director of Internet Policy
National Telecommunications and Information Administration
U.S. Department of Commerce
Washington, DC 20230

RE: SIIA Comments on Big Data and Consumer Privacy in the Internet Economy

Dear Mr. Morris,

On behalf of the Software & Information Industry Association (SIIA), thank you for the opportunity to provide comments as you consider "big data" developments and how they impact the proposed Consumer Privacy Bill of Rights (CPBR). These comments are provided in response to the National Telecommunications and Information Administration's (NTIA) request for public comment on "big data" and consumer privacy published on June 6, 2014.[1]

SIIA is the principal trade association for the software and digital information industry, representing more than 800 member companies. SIIA represents the industries that publish and distribute digital information, provide software applications and related Internet-based technology services. These industries are among the fastest-growing and most important industries of the U.S. and global economies, and they are critical drivers of data-driven innovation and digital trade.

SIIA produced a white paper in 2013 explaining data-driven innovation and how it presents tremendous economic and social value, capable of transforming the way we work, communicate, learn and live our lives.[2] In the paper, we explained the nature of this innovation, how it empowers enterprises and governments to benefit individuals, and we highlight how it is already enabling economic growth.

The report also identifies that it is crucial to maintain a public policy framework so that the protection of individual privacy can complement rather than thwart the natural evolution of new technologies. SIIA remains committed to working with technologists, privacy advocates and policy makers to foster the societal, governmental and business opportunities provided by data-driven innovation, while also meeting the challenge of protecting privacy. Data-driven innovation must be built on a foundation of good data stewardship, and SIIA supports the goal of government to encourage market participants to practice responsible use of data.

[1] Big Data and Consumer Privacy in the Internet Economy, 79 Fed. Reg. 109, 32714-32716, June. 6, 2014.
[2] SIIA, "Data-Driven Innovation, A Guide for Policymakers: Understanding and Enabling the Economic and Social Value of Data," April 2013.

As you consider revising the framework for the CPBR, SIIA submits the following set of broad considerations, as well as answers to some of the specific questions that you raise in the request for comment.

In general, SIIA believes that existing laws have continued to function quite effectively and provide a significant degree of protection, even in light of rapid technological innovation, increased data collection and analytics. Indeed, the current framework is well suited to enable a greater focus on responsible use of data and accountability to further enable data-driven innovation while protecting citizens.

I. Broad Considerations About the Intersection of Big Data and Privacy

(1) The emphasis should be placed on the responsible use of data and accountability, rather than unnecessary new limits on data collection and use.

As SIIA pointed out in our recent comments on big data[3], data and analytics have been around for quite some time. What is new with big data is the increasing capacity for enterprises and governments to more effectively gather this data, and to analyze and use it—from a variety of voluminous sources of structured and unstructured data, real-time and static—to innovate and improve the outcomes of everyday life. Entrepreneurs, established businesses, educational institutions and governments have increased abilities to put data to work to change the world for the better, applying these innovative abilities to everything from infrastructure, to financial services, education, healthcare, food production and consumer goods and services.

Larger data sets and more affordable analytical techniques increasingly enable greater insights and create greater value for organizations and individuals than were previously possible. One key novelty is that, in addition to finding answers to specific questions or challenges, big data analysis often allows insights that could not be anticipated empirically or theoretically before the analysis took place. Data analysis is no longer simply hypothesis testing. Instead, the data "speak" and tell the data scientists something they did not know before. This contrasts with historical practices of creating or coming more narrow sets of data with preset conclusions or objectives.

In addition to the broad societal benefits, and those obvious to governments and businesses, individuals and small businesses stand to benefit most from cost-effective, sophisticated data-powered tools and analytics systems they have never had access to before, to harness the power of their data to deliver practical benefits. This is often referred to as "democratization of data," where consumers and small businesses use data to make better decisions about everything from what they buy to how they plan for the future.[4] These decisions can be minor, such as customized services to

[3] SIIA comments to Mr. John Podesta, Counselor to the President, Executive Office of the President.
[4] Intuit, "The New Data Democracy: How Big Data Will Revolutionize the Lives of Small Businesses and Consumers," 2012.

an individual, or they can be major such as deciding where to go to college based on school evaluations or predictions of future career earnings.[5]

For these reasons, SIIA's overarching big data recommendation for policymakers, which certainly applies to the context of the CPBR, is that policies that have the effect of substantially curbing the collection, analysis and use of data threaten to stifle the nascent technological and economic revolution of big data and data-driven innovation before it can truly take hold.

Instead, SIIA believes that organizational policies that govern information collection, management and application can help to meet the various different legal, social and cultural requirements and expectations around the world. Indeed, accountability and good data stewardship—including transparency, data security and data quality—are critical to data-driven innovation reaching its full potential.

(2) To enable the benefits of big data and data-driven innovation, we must maintain an evolving view of privacy rights, balancing these with societal benefits.

Expectations surrounding the collection and processing of personal information are not purely personal. They reflect evolving social norms of the appropriate flow and use of information.[6] As technologies evolve to become instrumental in all facets of our lives, our experience and expectations of privacy also evolve.

In the past, particularly with the emphasis of Fair Information Practice Principles (FIPPs), policies have sought to increase individual choice and responsibility. However, the possibilities and benefits of data-driven innovation severely challenge the individualist paradigm of privacy, bringing about new social norms and expectations about the flow of information at a rapid pace. The voluntary sharing of information in exchange for benefits, including but not limited to personalized recommendations and customized services, reflects changing attitudes toward information sharing by millions of individuals. We highlight many of these examples in our white paper on data driven innovation, including the ability to improve public safety, health outcomes and empower small businesses to benefit from the power of data.[7]

Data-driven innovation is not unique to the private sector, but also increasingly implemented broadly by governments to improve the lives of citizens. Today, in cities around the United States, citizens can log on to integrated government websites to learn about neighborhood school performance and hospital wait times; and use that information to make important personal decisions, such as where they choose to move and where they seek treatment in an emergency. And, real-time weather forecasts, transit information, and health alerts, generated entirely from

[5] Daniel Castro & Travis Korte, Data Innovation 101: An Introduction to the Technologies and Policies Supporting Data-Driven Innovation, November 4, 2013.
[6] Helen Nissenbaum, Privacy in Context, Stanford University Press, 2010
[7] SIIA, Data-Driven Innovation, A Guide for Policymakers: Understanding and Enabling the Economic and Social Value of Data, April 2013.

government data—often through commercial smartphone applications—that further improve people's quality of life.[8]

Therefore, policy frameworks governing information sharing and use must remain sufficiently flexible to accommodate these evolutionary changes, recognizing that socially acceptable norms of information flows are evolving along with technology, balancing any constraints on information use against societal values such as public health, economic growth, the environment.

(3) Existing laws remain quite effective, and any new policies should build on the current risk-based framework, focused on preventing harm, where privacy and security are commensurate with the sensitivity of data.

Existing laws have, in many ways, continued to function very effectively and provide a significant degree of protection, even in light of rapid technological innovation, increased data collection and analytics. Together, the combination of various sectoral privacy laws such as the Health Information Portability and Accountability Act (HIPAA), Gramm-Leach-Bliley Act (GLBA), Fair Credit Reporting Act (FCRA), and Section 5 of the FTC Act, provides a framework where data privacy and security is commensurate with the sensitivity of data, where there is a heavy reliance on risk assessment and appropriate data uses by governments and entities.

For instance, SIIA recently published a white paper highlighting how the FCRA consumer protection framework is keeping pace with technological innovation to continue protecting consumers, and it provides a good model for policy in the age of data-driven innovation.[9] This approach is somewhat at variance from the standard notice and choice framework of privacy regulation. Instead of simply describing information use (that is, giving notice) and providing consumer choice, the "harm framework" seeks to identify the likely harms that the activities of these companies might cause, and then target any needed regulatory interventions to mitigate or reduce the risks of harm in a way that balances the costs and benefits involved.[10]

This model remains viable. Recent enforcement actions underscore that existing privacy laws are broad enough to encompass nascent technologies and business models, even though such technologies and business models may not have been contemplated when such privacy laws were enacted.[11]

[8] Ben Hecht, "Big Data Gets Personal in U.S. Cities," June 12, 2014.

[9] SIIA, "How the FCRA Protects the Public," December, 2013.

[10] J. Howard Beales, III & Timothy J. Muris, "Choice or Consequences: Protecting Privacy in Commercial Information" 75 U. Chi. L. Rev. 109 2008 pp. 109-120.

[11] The continued effectiveness of this model in the age of data-driven innovation was demonstrated in several recent cases, such as the FTC Spokeo case, the Filiquarian Publishing case and the Social Intelligence case**,** where it was determined that the law effectively covers entities that use the most advanced technology, including online data aggregation, social media and mobile apps for a wide range of eligibility contexts as the law was designed several decades ago.

II. Questions

(1) How can the CPBR, which is based on the Fair Information Practice Principles, support the innovations of big data while at the same time responding to its risks?

Fair Information Practice Principles (FIPPs) have provided guidelines for policymakers and data stewards regarding responsible information management practices for decades, throughout several major technology evolutions. However, as noted above, changing technological capabilities and shifting expectations of privacy have challenged the application of these principles in many cases. Therefore, crafting new public policies seeking to transform these principles into a set of legal "rights" is not likely to be effective.

For instance, it is widely recognized that there are significant limitation to the practical application of obtaining meaningful informed consent. This perspective was recognized by both the Administration's big data report, and that from the President's Council of Advisors on Science and Technology (PCAST). The Administration's report provided an elaborate recognition of the challenges and limitations of "notice and choice" framework, and therefore the need to focus a greater emphasis on a transparency and responsible use framework.[12] The report accepts that we may be required "to look closely at the notice and consent framework that has been a central pillar of how privacy practices have been organized for more than four decades."[13] The PCAST Report contains a similar recommendation that "policy attention should focus more on the actual uses of big data and less on its collection and analysis."[14]

New policies requiring affirmative consent could provide a substantial barrier to socially beneficial uses of information, not because people object to the collection or use, but because the process of obtaining, tracking and applying consent is itself too cumbersome or entirely impractical in many cases as data collection continues to increase by devices with small or no screens. Notice and choice will remain critical components in many specific or sensitive circumstances, but it cannot be the sole or even the primary mechanism for privacy protection in the age of big data, and it should certainly not be enshrined in broad one-size-fits-all policies. However, both policymakers and businesses together should continue to explore new models to improve transparency. SIIA was a leading participant in the recent NTIA multistakeholder discussion on mobile transparency, and we remain committed to working to improve voluntary practices in this area.[15]

Opportunities presented by data-driven innovation also challenge interpretations of "focused collection" or "data minimization," where data purpose specification and use limitation are overly rigid or prescriptive. The notion of data minimization is meant to protect individuals from privacy harms by collecting only the minimum amount of data and then destroying it as soon as possible.

[12] Executive Office of the President, "Big Data: Seizing Opportunities, Preserving Values," May 1, 2014, pp. 54-58.
[13] Ibid, p. 54.
[14] President's Council of Advisors on Science and Technology, "Big Data: A Technological Perspective," May 1, 2014, p. 49.
[15] SIIA Supports Mobile App Code of Conduct, July 25, 2013.

However, while the objective is laudable and the approach very practical in certain instances, there is a tension between this method of protecting privacy and the new capabilities presented by data-driven innovation, which thrive on enormous volumes of data and the discovery of novel, unanticipated connections within them.

Data-driven innovation is about maximizing data to identify new meaning and values among a wide range of seemingly unrelated data. Data minimization should not become a rigid construct. Rather it must continue to remain a key element of good data stewardship, which balances risk. For instance, there is no business need to store credit card security codes after a transaction has been processed, and saving such information creates substantial fraud risks. Where a reinterpreted data minimization principle would still dictate that such information not be retained, it would continue to allow data collection and retention for further analysis in the absence of demonstrated risk. Perhaps data minimization should come into play only when feasible and when there is substantial risk of harm and little likelihood of benefit from data collection.

In their recent Report on big data, PCAST reached a similar conclusion, questioning the value of such policies based on technological limitations, "there is little technical likelihood that 'a right to forget' or similar limits on retention could be meaningfully defined or enforced."[16] In determining that direct controls on collection are infeasible in most cases, and highlighting the impracticality minimization and deletion requirements, PCAST concludes that "attention to collection practices may help to reduce risk in some circumstances." And that best practices such as tracking provenance, auditing access and use, and continuous monitoring and control could possibly arise from partnership between government and industry.[17]

SIIA continues to believe that the combination of privacy by design techniques and adherence to a set of responsible data principles can create an effective framework that balances privacy with innovation and accounting appropriately for risk, without the creation of a broad set of legal "rights" based on FIPPs.[18]

(2-3) Should any of the specific elements of the CPBR be clarified or modified to accommodate the benefits or risks of big data? Should a "responsible use framework" be used to address some of the challenges posed by big data?

The Administration's Big Data Report establishes what could be interpreted as a *data fairness principle*, stating broadly that, "it is the responsibility of government to ensure that transformative technologies are used fairly and employed in all areas where they can achieve public good."[19] This

[16] President's Council of Advisors on Science and Technology, "Big Data: A Technological Perspective," May 1, 2014, p. 48.
[17] Ibid.
[18] SIIA, "Data-Driven Innovation, A Guide for Policymakers: Understanding and Enabling the Economic and Social Value of Data," April 2013, pp. 16-18.
[19] Executive Office of the President, "Big Data: Seizing Opportunities, Preserving Values," May 1, 2014, p. 48.

assertion is very ambitious, and it is not entirely consistent with the narrower concept of a "responsible use framework," also discussed in Section 5 of the report.

SIIA shares the sentiment in the Big Data Report that new technologies ought not give rise to new forms of discrimination, but we are not convinced that the existence of these new forms of discrimination have been demonstrated. So far they seem to be hypothetical, even identified by the Administration's big data review as the, "…significant finding of our review was the the *potential* for big data analytics to lead to discriminatory outcomes and to circumvent longstanding civil rights protections in housing, employment, credit, and the consumer marketplace." [emphasis added][20]

Moreover, existing laws against discriminatory uses of information provides strong and robust protections and appear to already cover these hypothetical examples of discriminatory big data use as a need for new policies. To the extent any new policies are determined to be necessary, we recommend that they be narrowly tailored, building off of existing policies where gaps are identified that lead to identifiable harms, rather than a broad framework that seeks to maintain a preset definition of "fairness" or raise new barriers to data collection and use in an attempt to prevent harms that "could" arise.

Recognition of this perspective has been increasing substantially. For instance, Benjamin Wittes also suggests that the possibility of harm is not enough to warrant regulation, saying, "…for a privacy claim to be cognizable as a problem warranting public policy attention, there must be some asserted harm…" which he defines as "the malicious, reckless, negligent, or unjustified handling, collection, or use of a person's data in a fashion adverse to that person's interests…"[21] Adam Thierer also notes the defect in "precautionary" thinking that recommends regulatory action solely on the basis of possible harms, saying, "It is not enough to claim, 'Well, it *could* happen!'"[22]

As stated above, SIIA agrees that the biggest challenge posed by a CPBR is that it creates a construct too rigid, rather than focusing on use, mitigating risk and seeking to prevent harm. Given the challenges of a FIPPs-based approach to developing a CPBR, which would rely heavily on increased government regulation for addressing the principles, SIIA believes that current law effectively provides for an accountability framework where responsibility for protecting consumers can be shifted from the data subject to the user.

Data collection is not in itself harmful, and in an age of ubiquitous data collection any attempt to impose controls on data collection improperly puts the entire burden of regulation on data subjects. The need for this realignment of responsibilities was highlighted several years ago by Daniel Weitzner and his colleagues who said:

> Consumers should not have to agree in advance to complex policies with unpredictable outcomes. Instead, they should be confident that there will be redress if they are harmed by

[20] John Podesta, "Findings of the Big Data and Privacy Working Group Review," May 1, 2014.
[21] Benjamin Wittes, "Databuse: Digital Privacy and the Mosaic," Brookings Institution, April 1, 2011, p. 17.
[22] Adam Thierer, "Permissionless Innovation," Mercatus Center, George Mason University, 2014, p. 31.

> improper use of the information they provide, and otherwise they should not have to think about this at all. [23]

In addition, a number of scholars have stressed the importance of internal accountability and data stewardships, principles which SIIA strongly supports. For example, there is the interesting thought experiment of consumer subject review boards suggested by Ryan Calo.[24] The idea is that companies should appoint a small group of employees with different backgrounds to assess data projects involving consumers. Victor Mayer-Schonberger and Kenneth Cukier have a similar suggestion of an internal ombudsman (an algorithmist) who would internally vet projects.[25] Institutional reforms that provide more internal accountability might be one way to implement an accountability framework in general and in particular for the development and use of consumer scores, but a new regulatory framework is not warranted at this time.

Rather, governments are best positioned to provide guidance and further incentivize companies to devote internal resources towards accountability and prevention of harm. Voluntary best practices and codes of conduct continue to be very effective in this regard, therefore SIIA continues to be a strong supporter of the NTIA multistakeholder process devoted to the development of voluntary best practices. This process might be well served to explore approaches to encourage accountability and good data stewardship.

(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 CPBR's "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?

Individual control and respect for context are both very important principles of good data stewardship. That said, both are difficult to apply broadly within a set of "rights" for citizens.

The concept of "respect for context" or that that privacy requirements should apply when the business context calls for it, but not when the information practices in question are commonly accepted business practices, can be an effective principle, as a first step. In exploring the evolving privacy framework focused on data use, the recent Big Data Report reiterates the Administration's support for "respect for context" concept established in the initial CPBR, characterized as a "no surprises" rule. For instance, "data collected in a consumer context could not suddenly be used in an employment one."[26] The report also highlights that "technological developments support this shift toward a focus on use," noting that "advanced data-tagging schemes can encode details about

[23] Daniel J. Weitzner, et al., "Information Accountability," Computer Sci. & Artificial Intelligence Laboratory Technical Report MIT-CSAIL-TR-2007-034, 2007.
[24] Ryan Calo, "Consumer Subject Review Boards: A Thought Experiment," 66 Stan. L. Rev. Online 97, September 2013.
[25] Mayer-Schönberger & Cukier, "Big Data," pp. 181-182.
[26] Executive Office of the President, "Big Data: Seizing Opportunities, Preserving Values," May 1, 2014, p. 56.

the context of collection and uses of the data already granted by the user, so that information about permissive uses travels along with the data wherever it goes."[27]

There are instances where secondary use of data might be surprising and potentially harmful, such as the employment example cited. However, potential harms posed would already be covered by current law, such as the FCRA. A blanket prohibition on unexpected or out-of-context uses of information applied as part of a broad policy framework could unnecessarily require privacy restrictions even when there is no risk of consumer harm.

Additionally, there are significant challenges in applying individual control requirements in an era of big data. That is, given the dynamic and evolving nature of data in the big data context, it is not practical to think that access and correction could be effectively implemented across the board. SIIA is particularly concerned with prescriptive mandates that "data brokers" providing marketing services could, and should, effectively allow for consumers to correct their personal information for marketing purposes.[28] In addition to the challenges of implementing such "control" in a dynamic big data environment, there would be substantial new challenges posed by the nature of rapidly evolving data sets and authentication. In this case the desired "privacy" outcome could potentially create greater challenges and risks than currently exist today.

But that is not to say there is no use for individual control, which very much will continue to be a fundamental pillar or good data stewardship in many circumstances. For instance, this is the approach behind existing regulatory regimes such as FCRA and the FTC's section 5 unfairness authorities.

There has been considerable assessment by the FTC of the need for increased individual control in the case of "alternative" or "consumer scoring," but the existence of sufficient risk or harm in this area has not been sufficiently identified. In recent comments to the FTC, SIIA concluded that the current statutory and regulatory framework seems to be adequate for addressing the issues raised by the use of predictive analytics in general and the use of consumer scores as described in the Commission's March 19 workshop. However, the FTC should monitor the marketplace (1) to take strong and effective enforcement measures against firms that violate current statutory or regulatory constraints and (2) to ascertain whether there are business practices that could lead to consumer harm, but are not addressed adequately within the current framework. A general workshop exploring the concept of consumer harm in more detail might be helpful as well, since the workshop revealed substantial differences in views regarding which business practices constituted consumer harm.[29]

At this time, the notion of individual control being applied as a broad requirement on less sensitive data would only seek to stifle innovative uses without providing significant new protections for consumers.

[27] Ibid.
[28] U.S. Federal Trade Commission, "Data Brokers: A call for Transparency and Accountability," May 2014.
[29] SIIA Comments to the U.S. Federal Trade Commission,

(10-11) How significant are the privacy risks associated with "data fusion?" How significant are the privacy risks posed by re-identification of de-identified data? How can re-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? 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 technologies?

Some of the most important outcomes of big data and data-driven innovation do not rely on personally identifiable information. Indeed, a lot of high-value analytics result from non-personal information, by simply looking at aggregate customer data. Often, valuable analysis can be done where any particular individual or customer is no more than an arbitrary, non-traceable number. It is very common that, even if personal information is collected, companies will take steps to immediately de-identify the data in a way that does not affect its value or utility for accomplishing important public and social objectives. In fact, many data scientists say they do not want to use identifiable data because they find it too specific to determine meaningful insights and group characteristics.

Two years ago, the FTC's 2012 privacy report proposed a new definition of what should be considered de-identified data, and thus outside the scope of what is considered personally identifiable information. The report effectively called for reduced privacy regulation when data meets a three-factor test: "(1) a given data set is not reasonably identifiable; (2) the company publicly commits not to re-identify it, and (3) the company requires any downstream users of the data to keep it in de-identified form."[30]

Notably, the FTC definition relies heavily on organizational controls. The company commits not to re-identify data and it must impose that commitment on its downstream users. The commitment not to re-identify data fits well into the FTC's Section 5 authority to prohibit deceptive trade practices – once a company has promised not to re-identify data, it would presumably be deceptive under Section 5 to break that promise.

The FTC also notes that "what qualifies as a reasonable level of justified confidence depends upon the particular circumstances, including the available methods and technologies. In addition, the nature of the data at issue and the purposes for which it will be used are also relevant."[31]

By comparison, the technical component of the FTC's proposal appears less strict than for HIPAA. The FTC requires only that "a given data set is not reasonably identifiable," in contrast to the HIPAA requirement of a "very small risk that the information could be used, alone or in combination with other reasonably available information," to re-identify the individual. In considering

[30] U.S. Federal Trade Commission, "Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers," 2012.
[31] Ibid.

reasonableness, the FTC standard applies to a wide range of PII held by commercial organizations, rather than only to sensitive health information.

SIIA at that time strongly supported that conclusion, and we continue to support the conclusion where de-identification techniques are adequately applied, commensurate with the sensitivity of data, data sharing practices and the existence of other controls.

Of course, de-identification techniques have come under significant scrutiny over the last couple years. Most recently, PCAST recognizes the popularity of this practice, but goes on to question the value of de-identification and anonymization. In general, the report declares that it is "increasingly easy to defeat anonymization by the very techniques that are being developed for many legitimate applications of big data," and that "in general, as the size and diversity of available data grows, the likelihood of being able to re-identify individuals grows substantially."[32] In conclusion, PCAST characterizes anonymization as only a "somewhat useful added safeguard," but not sufficiently robust to be a useful basis for policy.

This conclusion is accurate in some respects. First, it identifies the challenges of perfect de-identification, and the shortfalls of policy requirements around this approach. However, the conclusion in many ways understates the continued value and effectiveness of de-identification, which continues to be a legitimate basis for reduced privacy concerns and regulation. Even if not perfect in many cases, de-identification can allow for robust privacy protection, safely enabling innovative and societally beneficial purposes without posing significant risk of re-identification.

A recent report provides a strong defense of the value of de-identification, highlighting that the risk of re-identification of individuals from properly de-identified data is significantly lower than indicated by many commentators. The report provides that the continued lack of trust in de-identification and the myths about the ease of re-identification may make data custodians less inclined to provide researchers with access to much needed information, even if it has been strongly de-identified, or worse, to believe that de-identification is a waste of time and therefore.[33] Either way, the risks of understating the value of de-identification are likely to be a self-fulfilling prophecy, serving to delegitimize a very effective and practical approach that has protected sensitive data for many years, and can continue to do so in many contexts, even in an era of big data.

SIIA urges policymakers to encourage de-identification as a way to balance the needs of data-driven innovation and privacy protection, but to avoid broad mandates to this end. Policymakers should also seek to support the development of strong tools, training and best practices so that these techniques may be more widely adopted. Examples include a governance structure in place to enable organizations to continually assess the overall quality of their de-identified datasets to ensure that their utility remains high, and the risk of re-identification sufficiently low.[34]

[32] President's Council of Advisors on Science and Technology, "Big Data: A Technological Perspective," May 1, 2014, p. 38-39.
[33] Cavoukian, Ann, and Castro, Daniel, "Big data Innovation, Setting the Record Straight: De-identification Does Work," June 16, 2014.
[34] Ibid, p. 9.

Further, an additional caution is that if information that is not individually identifiable comes under full remit of privacy laws based on a possibility of it being linked to an individual at some point in time through some conceivable method—no matter how unlikely—this could not only prohibit many beneficial uses and benefits of data-driven innovation, but it could also destroy the incentive to de-identify the data. The FTC avoided this mistake in its practical approach to de-identification.[35]

(12) Should the CPBR address the risk of discriminatory effects resulting from automated decision processes using personal data, and if so, how? Should big data analytics be accompanied by assessments of the potential discriminatory impacts on protected classes?

As stated above, given the tremendous benefits offered by data-driven innovation, we shouldn't adopt broad policies that seek to address harms—in this case potential for discrimination—that merely *could* happen, but rather policies should be focused on identifying actual harms occurring today but not addressed due to gaps in current law. If in the future there is compelling evidence that additional consumer protections are needed, then new protections should be undertaken at the stage of data usage or implementation, rather than at the early stages of data collection or analysis.[36]

Moreover, no new privacy requirements should be extended to external service providers or industry infrastructure providers. These entities perform a necessary role in the industry, but they do not themselves collect or direct the use of consumer information. They usually serve merely as service providers to other entities that actually make use of that data. As such, they are removed from decision making regarding information collection and use and should be immune from any new privacy requirements that might be contemplated.

The PCAST report explores this topic at length, highlighting that "an increasing fraction of privacy issues will surface only with the application of data analytics," and that "many privacy challenges will arise from the analysis of data collected unintentionally that was not, at the time of collection, targeted at any particular individual or even group of individuals," as data science gains power to combine and comb data from a wide variety of sources. Therefore, PCAST contemplates whether it might be feasible to introduce regulation at the "moment of particularization" of data about an individual, or when this is done for some minimum number of individuals concurrently.[37]

However, there are a number of reasons why this approach would not be effective. As PCAST also highlights, such entities would be difficult to regulate because their actions do not directly touch the individual via either collection or use and may have no external visibility. Also, PCAST points out that and any regulation would need to be accompanied by requirements for tracking provenance,

[35] U.S. Federal Trade Commission, "Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers," 2012.
[36] T.Z. Zarsky, "Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society," 56 Me. L. Rev. 13, 2004, p. 49.
[37] President's Council of Advisors on Science and Technology, "Big Data: A Technological Perspective," May 1, 2014, p. 48.

auditing access and use, and using security measures at all stages of the evolution of data, and for providing transparency, and/or notification, at the moment of particularization.

SIIA concurs with PCAST's conclusion with respect to the application of any new regulation. After assessing both the policy and technical realities in this area, PCAST reaches the following conclusion:

> It is not, however, the mere development of a product of analysis that can cause adverse consequences. Those occur only with its actual use, whether in commerce, by government, by the press, or by individuals. This seems the most technically feasible place to apply regulation going forward, focusing at the locus where harm can be produced, not far upstream from where it may barely (if at all) be identifiable. [38]

Impacts of big data on underserved groups are not one-sided, as the use of data and data analytics can uncover and help prevent disparate impacts on protected classes. That said, SIIA would be concerned about a blanket requirement for all uses of data to be accompanied by a disparate impact analysis. Such assessments are often quite resource intensive and cannot be done quickly, and could therefore only be reasonably justified in cases which are narrowly tailored to specific uses of data where significant risk is shown to exist.

The FTC's assessment of the use of credit insurance scores mandated by Congress took a team of economists a year to complete. The President's big data report focused on the need for federal agencies responsible for discrimination laws to devote increased resources to the use of data to uncover patterns of discrimination.

Policymakers must balance the benefits with potential challenges, recognizing the ability of data to improve access to healthcare, credit and even help ensure access to government benefits and tax credits to those most in need. It should be a priority for policymakers to encourage market participants to use data for these and other beneficial uses, and to take reasonable steps to use data and data analytics to further understand potential discrimination.

III. Conclusion

Again, thank you for the opportunity to comment as you consider these important issues. If you have further questions or would like to discuss, please do not hesitate to contact David LeDuc, SIIA's Senior Director for Public policy, at dleduc@siia.net or (202) 789-4443.

Sincerely yours,

Ken Wasch

Ken Wasch
President

[38] Ibid, p. 49.