Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399

Tel 425 882 8080
Fax 425 936 7329
http://www.microsoft.com/

August 5, 2014

Mr. John Morris
National Telecommunications and Information Administration
U.S. Department of Commerce
1401 Constitution Avenue NW, Room 4725
Attn: Privacy RFC 2014
Washington, DC 20230

Re: Request for Comment: Big Data and Consumer Privacy in the Internet Economy

Dear Mr. Morris:

Microsoft welcomes the thorough and thoughtful results of the White House's big data review, including the Big Data Report[1] and the PCAST Report,[2] and we are pleased to respond to the Request for Comment. The subject is of keen interest to Microsoft because we are building mobile and cloud technologies that enable us to deliver innovative services to our customers in part through analysis of large amounts of data, and we offer technologies that enable our business customers to do the same.

As the Big Data Report recognizes, big data offers the opportunity to grow the U.S. economy, improve health and education, and make our nation safer, more energy efficient, and more productive. But at the same time, growing concerns about potential risks to privacy from big data threaten to constrain its potential.

Microsoft understands this dynamic well. We work hard every day to develop technologies and processes that deliver robust, easy-to-use privacy protections for our customers. At the same time, Microsoft, and the technologies we deliver—such as our cloud platform Microsoft Azure—are at the forefront of the big data revolution. Experience has taught us that strong data protection practices are not the antithesis of innovative data usage. Rather, privacy and big data can and must go hand-in-hand.

The United States is well placed to take a leadership position on privacy and big data. But it needs to move quickly. Microsoft encourages the Department of Commerce and the Administration more

[1] Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (May 2014), *available at: http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may*.
[2] Executive Office of the President, President's Council of Advisors on Science and Technology, *Report to the President, Big Data and Privacy: A Technological Perspective* (May 1, 2014), *available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data*.

broadly to push for passage of comprehensive federal privacy legislation based on the principles in the Consumer Privacy Bill of Rights. The rise of big data makes this a pressing issue.

The Need for Comprehensive Federal Privacy Legislation

The nation should move forward on privacy legislation now. Strong, comprehensive federal privacy legislation could establish a framework that enables all players to harness the potential of big data while respecting the privacy rights of those whose information contributes to the data. Microsoft has supported privacy legislation at the federal level for many years, and the rise of big data only increases the need for action.

Our key trading partners understand the urgent need for privacy legislation that is suited to the big data era. The European Union is in the midst of reforming its privacy laws, developing new rules that will govern privacy in all 28 EU countries. In Asia, Japan recently issued terms of reference for a new privacy bill to replace its 2003 law, and Korea and Singapore have new laws in place this year. Markets throughout Latin America are moving to adopt and update privacy frameworks. The United States should not stand still.

Without new privacy legislation, U.S. companies will find themselves increasingly disadvantaged compared to foreign providers that will compete against U.S. companies in their home and other jurisdictions based on more protective privacy regimes. Over time, absent sound rules of the road, it will likely become harder for U.S. companies to keep the trust of consumers worldwide. Already, some customers for cloud services in foreign markets are turning towards local solutions instead of U.S. providers, precisely because they (and their regulators) do not trust to the sufficiency of U.S. privacy laws.

This lack of trust also may be compounded over time as countries adopt new privacy frameworks that—following in the footsteps of the European Union—restrict data flows to the United States out of concern that data will not be robustly protected here. The U.S.-based big data industry will not thrive if it is cut off from global data.

The adoption of a comprehensive U.S. privacy law may, conversely, encourage the flow of data to the United States, triggering increased physical data center infrastructure and generating more big data-focused jobs and growth here at home. A comprehensive U.S. regime may also act as a counterpoint to more restrictive third-country proposals, inspiring countries to adopt a less protectionist view of privacy, and encouraging the free flow of data globally—to the benefit of businesses and their customers both in the United States and abroad.

Microsoft also encourages U.S. regulators to continue their constructive ongoing dialogue with international privacy authorities, particularly those in the European Union. International collaboration with the aim of developing interoperable privacy frameworks is essential given expanding data flows across national boundaries and the cross-border nature of big data business.

Adapting Notice and Consent for the 21st Century

The NTIA's Request for Comment asks about the practical limits of the "notice and consent" model, which is the basis of many privacy laws worldwide. This is an important subject. If we do not get notice and consent right, we will not fully benefit from big data analysis that bears upon people.

A simple notice and consent model made good sense in an era of one-to-one data collection. But as technology has advanced in the decades since the notice and consent model first was developed, it has become harder to provide complete notice of all the data that is collected and to seek detailed consent for every use of that data.

Data collection is ubiquitous today, and data is often gathered without the intervention of the individual. That makes both giving notice and getting consent to use personal data more difficult through the traditional means of privacy policies and one-time consents to agreements. Moreover, users often find over-notification disruptive to their online experience.

At the same time, it is essential that individuals both understand and have a meaningful way to participate in and control how their data is used. Absent that control, individuals will lose trust in online services. We thus should not abandon notice and consent.

Instead, notice and consent must be adapted—and strengthened—for the 21st century. The Consumer Privacy Bill of Rights does that. By focusing on transparency and individual control, the Consumer Privacy Bill of Rights extends the concepts of notice and consent. Mere "notice" implies a one-time, one-way disclosure, reflecting a traditional approach, often through ever-lengthening privacy policies. "Transparency" extends that concept through continuous activity involving both the data collector and the individual. Privacy notices are part of that dynamic, but they are not the only part. Companies that collect personal data must be forthcoming about how they will use it and do so in ways that go beyond mere static privacy policies. Companies should make this information readily available to individuals through online settings and customer support—including "just in time" notices as the Request for Comment notes.

Similarly, practices for obtaining consent in today's big data world should be strengthened. The Consumer Privacy Bill of Rights recognizes this in its individual control principle. Under that principle, the individual does not cede power to a data collector through a one-time consent. Instead, the individual and the data collector remain in a relationship that may change over time, and one in which the individual remains actively engaged. This has special benefits for the use of big data, as one of the attributes of big data is that it creates insights that researchers may not at first expect. Maintaining a continuous relationship of transparency and providing individuals meaningful control over their personal data allows the relationship among big data, the data collector and the individual to evolve.

This point is important: big data *depends* on the evolution of data usage over time. Data collected for one purpose—for example, customer relations management—may over time help companies to reveal commercial trends and new opportunities for growth. The value of big data is therefore closely related to helpful and appropriate additional uses of data. Among the challenges with the traditional model of

consent is that, once consent is collected, the data usage purposes covered by the consent can be "locked in" and not change further without a difficult attempt by a business to seek additional or expanded consent. This makes it difficult to extract new insights from data sets about people.

A more dynamic relationship, based on transparency, would help businesses unlock greater value through additional usages of data, speeding U.S. big data efforts. That said, permissions for businesses to use data for additional purposes should not be without limit—constraints should be imposed by law to ensure that such uses are kept proportionate, reasonable, and in line with consumer expectations. Additional safeguards, such as support in the law for de-identification wherever feasible, also should be considered.

Microsoft also supports implementing in legislation the third principle from the Consumer Privacy Bill of Rights, respect for context. More transparency and greater control should be required when companies use data for their own benefit or for the benefit of others, outside the context expected by the individual providing his or her data. Too often, companies' practices today do the opposite. Companies often are the most transparent about *obvious* uses which are unlikely to concern customers—for example, organizations will describe at length how online complaint forms are used to contact complainants and resolve complaints—but can be much less transparent about non-obvious data uses, such as online tracking or scanning of emails to target advertising based on users' behaviors.

Big data increasingly drives companies to seek new data collection methods and data uses. On the whole, this can deliver growth and more efficiency for businesses, and should be encouraged. But, to the extent such collection or use goes beyond what consumers would expect, enhanced notice and transparency obligations should take effect.

These requirements will not be overly burdensome for big data businesses. Microsoft has found that by employing a "privacy-by-design" approach when engineering new products and services, we can offer a high level of transparency and user control. And we continue to develop more granular tools for users to exercise choice over how their data is used.

Technological Challenges Created by Big Data

As the PCAST Report notes, big data is unlike traditional personal information in that it can be analyzed to identify individuals indirectly, including re-identifying individuals in data sets that were thought to be anonymous. To address the risk of re-identification, future privacy legislation should define the scope of data subject to its provisions broadly. The Consumer Privacy Bill of Rights applies to personal data, defined as "any data, including aggregations of data, which is linkable to a specific individual." This definition matches definitions worldwide, and if properly interpreted will ensure that companies will be held accountable for their use of data that may be *identifiable to* individuals even if the data does not *directly identify* individuals.

At the same time, regulations should encourage those who hold data to do what they can to de-identify data when it is in use. By reducing the linkability of a data set to the original data subject, de-identification enhances the privacy and security of personal data. This does not mean, however, that

U.S. regulations should rigidly define concepts of "anonymous" or "pseudonymous data" by specifying data types that must be removed from a data set. This approach will prove inadequate; removing or replacing some types of data will be sufficient to ensure that the data set cannot be relinked to a specific individual in some contexts, but not in others. To give an example, if only a small number of people have a rare disease in a given city, it may not be sufficient to replace the name, date of birth, and address from the data set with a code, because it may be possible to identify an individual from a data set on the basis of a combination of other, non-obvious identifiers, such as the type of rare disease, gender and city of residence.

Because the assessment of whether a data set is sufficiently de-identified is context-driven, mandating the optimal techniques for de-identification would be counterproductive, as the effectiveness of a particular technique depends on the specific data set and the context where it is deployed. Moreover, given how rapidly de-identification techniques are evolving, any law that attempts to mandate particular techniques may quickly become obsolete. For these reasons, the focus of any rules in this space should be less on process and more on the desired outcome of that process.

Regulations also should prohibit using data to discriminate against individuals illegally, even when the individuals are not identified. The misidentification or misclassification of individuals can be as much of a privacy harm as identification—and the law should reflect this.

The Need for Robust Regulatory Oversight

The NTIA's Request for Comment explores the role of regulators and enforcement in a new U.S. data protection framework. While the Consumer Privacy Bill of Rights could be codified in a statute today, properly addressing the issues in the Big Data and PCAST Reports would require dedicated regulatory oversight.

The need for regulatory oversight is not limited to technical issues raised by big data. The NTIA's Request for Comment identifies one area where regulatory oversight could be particularly helpful: improper discrimination against individuals and groups. The Consumer Privacy Bill of Rights' access and accuracy principle touches on this, but that principle and the Consumer Privacy Bill of Rights may need to be extended to address problems of discrimination. The Big Data Report identifies instances where people have multiple surnames or where a woman adopts her husband's surname resulting in error rates. Database holders must be responsible for avoiding such errors under the accuracy principle. More subtle problems will occur when the data collected from individuals is biased in ways that only become visible indirectly through big data analysis. Identifying these problems and rectifying them will require sustained effort, in the context of laws and regulatory oversight. Big data analysis itself may help identify and confirm instances of discrimination. Regulatory cooperation between anti-discrimination regulators and privacy regulators could help make that a reality.

Unlocking the Value of Data Through Data Fusion

The NTIA's Request for Comment asks about the risks involved in data fusion. Data fusion, also known as data combination, involves the linking of multiple data sets (or discrete points of data about an

individual). This technique is critical to effective big data analysis. Done carelessly and without appropriate protections in place, data combination can increase privacy risks, because, for example, hackers gaining access to a database can obtain more data than they would have if data had not been combined.

Data combination privacy risks can be responsibly managed, however. In particular, organizations can avoid combining data more than is necessary to achieve big data aims; can de-combine data where appropriate; can de-identify combined data sets; can adopt enhanced security measures proportionate to the increased risk of combined data; and can limit the retention of combined data. These steps can mitigate privacy risks and enable businesses to productively employ big data techniques without endangering individuals' privacy.

Data combination is not only vital to big data techniques, but also to the increasing trend toward delivering personalized Web services to consumers. Personalization can only occur when increasing amounts of data are linked to specific user accounts. For example, email accounts are typically linked with contacts and calendar entries to make it easier for users to send mail to frequent contacts and schedule appointments. Any reform of privacy laws should recognize this important benefit of data combination. Microsoft supports reasonable safeguards in relation to data combination, but regulatory scrutiny should focus on the *use* of combined data, rather than the mere act of combination.

Developing Harmonized Breach Notice Obligations

The NTIA's Request for Comment asks about other approaches to promoting privacy in a world of big data. Consistent with the recommendations in the Big Data Report, Microsoft supports adoption of a pre-emptive federal breach notice standard. Customers will place greater trust in enterprises if they know that they will be advised of breaches that pose real threats to their privacy.

The current patchwork of state data breach rules, while providing some protection, creates compliance challenges for organizations handling data from customers nationwide as notification procedures, and the required contents of those notifications, vary and, in some cases, conflict. Moreover, many data breaches affect consumers across multiple states—making consistent rules at the federal level critical. Federal level legislation should be technology neutral, again to encourage consistency of consumer expectations and to enhance trust. Any such law should also take context and the privacy impact of each breach into account, in order to avoid "de-sensitizing" consumers with too many data breach notifications. Accordingly, aggregated and de-identified data sets (often used in big data applications) should not be subject to mandatory breach rules. Any national breach notification law should provide organizations with sufficient flexibility to determine the most effective way to notify consumers.

Government Access to Big Data

The Request for Comment focuses primarily on how the private sector can responsibly handle big data. But to fully address the protection of consumer privacy in the era of big data, legislative reform must also address how law enforcement, intelligence agencies, and other government agencies access and handle personal information. This is perhaps the biggest challenge to building public confidence in the

cloud and other emerging technologies that rely on big data. We cannot comprehensively address individual privacy without evaluating the limits on government data collection and surveillance.

As a recent survey released by Microsoft demonstrates, consumers want reasonable limits to be imposed on the government's ability to access data that they store in the cloud or on computing devices.[3] Indeed, when the Supreme Court in June unanimously required warrants for searches of cell phones in *Riley v. California*, it based its decision largely on the fact that people store highly personal information on their phones and in the cloud, and that they have a high expectation of privacy in that data.

Quite simply, people will not use technology that they do not trust. And they will not trust the cloud if the government has easy and insufficiently controlled access to the data that is stored in it. The government should at a minimum take the following steps to build confidence in the cloud:

- Update the Electronic Communications Privacy Act to address changes in technology.[4]
- Reform the Foreign Intelligence Surveillance Court to ensure that its proceedings are the product of the adversarial process that is the hallmark of our judicial system.
- Commit not to hack data centers or cables.
- Increase transparency about the amount and types of information collected through intelligence surveillance.
- End bulk collection of data of telephone records.
- Work with our international allies to improve the Mutual Legal Assistance Treaty process, and use that process to obtain digital evidence stored overseas, rather than using unilateral processes.

Conclusion

These issues, and the other issues identified in the Request for Comment, the Big Data Report and the PCAST Report are solvable. Big data holds great promise for technology, for the economy and for people. Comprehensive privacy legislation based on the Consumer Privacy Bill of Rights can and should be part of unlocking that promise. Microsoft supports such legislation and welcomes the work of the Administration in its efforts to explore these issues and build a policy framework to support privacy in the era of big data.

Yours sincerely,

David A. Heiner
Vice President & Deputy General Counsel, Legal and Corporate Affairs
Microsoft Corporation

[3] The survey is available at http://blogs.technet.com/b/microsoft_on_the_issues/archive/2014/07/16/digital-common-sense-new-survey-shows-americans-want-a-better-privacy-balance.aspx.
[4] For details, see the principles advanced by the Digital Due Process Coalition at http://digitaldueprocess.org.