mozilla

March 31, 2014

Office of Science and Technology Policy
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, DC 20504

RE: Office of Science and Technology Policy Request for Information: Big Data

To Whom It May Concern:

Mozilla submits these comments in response to the March 4, 2014 Notice of Request for Information on Big Data.

These comments emphasize that we as a multi-stakeholder Internet society are in the early stages of understanding big data, counseling deliberation and patience; note the complexities associated with surveillance and accentuated by big, global data; and suggest ambitious research and development to advance big data opportunities.

We greatly appreciate the Administration's efforts to lead a multi-stakeholder process to explore big data issues through this request for information and the three workshops.

On behalf of Mozilla, we thank you for the opportunity to comment on this request for information. Please do not hesitate to contact us with questions or for additional input.

Respectfully Submitted,

/s/

Alexander Fowler, Global Privacy and Public Policy Lead
M. Chris Riley, Senior Policy Engineer

Mozilla
2 Harrison St
San Francisco, CA 94105

mozilla

Comments on the Office of Science and Technology Policy Request for Information: Big Data

Prepared by Mozilla and Submitted on May 31, 2014

I. Introduction

Mozilla appreciates the opportunity to provide input to the Office of Science and Technology Policy (OSTP) for its "Big Data" review process. The expansion of massive data gathering, storage, and processing capabilities, together with the reduction in cost, are having a significant impact on society's collective understanding of the proper norms and approaches in this space. This OSTP process can build on top of existing academic, industry, and civil society efforts to catalyze next steps forward on research and multi-stakeholder discussion of these issues.

Mozilla is a global community of people working together since 1998 to build a better Internet. As a non-profit organization, we are dedicated to promoting openness, innovation, and opportunity online. Mozilla and its contributors make technologies for consumers and developers, including the Firefox web browser and Firefox OS phone used by more than half a billion people worldwide. As a core principle, we believe that the Internet, as the most significant social and technological development of our time, is a precious public resource that must be improved and protected.

Privacy and security are important considerations for Mozilla. They are embraced in the products and services we create, and derive from a core belief that consumers should have the ability to maintain control over their entire web experience, including how their information is collected, used and shared with other parties. We strive to ensure privacy and security innovations support consumers in their everyday activities whether they are sharing information, conducting commercial transactions, engaging in social activities, or browsing the web.

At the outset, it is important to establish improved trust as the goal of big data policy. This process is not about best practices to extract maximum revenue from big data sets, or disrupting industries. Instead, the challenge is to address the policy and normative concerns that arise from big data, and to understand how frameworks for privacy and trust developed for a different world should extend and adapt to this one. Within this challenge, the fundamental risk associated with big data policy is trust in the global Internet and information ecosystems, and whether the world's people and businesses will continue to participate in these ecosystems and realize their benefits for social, economic, and political activity.

These comments will raise three main themes responsive to the questions articulated by OSTP:

1. First, we as a multi-stakeholder community working on, and with, big data are very early in the process of understanding how everything works, including how to apply the norms developed for an earlier data and privacy world, and where both fuzzy and bright lines should be drawn around data handling practices that support innovation and growth, on the one hand, while preserving user control and driving public benefits.
2. Second, we must tackle head-on the heightened sensitivities and trust risks associated with government access to personal data, or we will not have a strong, internally cohesive, collaborative community to tackle these issues.
3. Finally, we should think "big." Big data presents big problems, but there are also big opportunities, and we should embrace them, not disregard ambitious or long-term solutions.

II. Policy Challenges that Arise from or are Amplified by Big Data

This section highlights a few key issues that, while not exhaustive, we think are the most important for the OSTP to grapple with to start:

A. the complex nature of "personal" information,
B. the significance of data portability,
C. growing concerns over balkanization of data systems, and
D. the inherently unique nature of government collection of and access to data.

Across these issues, the basic concepts of the Consumer Privacy Bill of Rights persist, including the value of control and transparency and the importance of context. Trust remains the ultimate normative goal. Yet, these all carry somewhat different meanings and implications in the big data world, and we don't yet have a full understanding of how to apply and implement them.

A. Properly Defining "Personal" Data

Responsive to Question 1: What are the public policy implications of the collection, storage, analysis, and use of big data? For example, do the current U.S. policy framework and privacy proposals for protecting consumer privacy and government use of data adequately address issues raised by big data analytics?

One fundamental challenge to getting data policy right is changing the binary concept of "personal" and "not personal".[1] Big data policy will struggle to be accurate if it rests on a broken categorization of unit data. This is responsive to Question 1, in that big data can amplify and extend the policy problems that arise from a misconceptualization of unit data, and in that big data creates new nuances of personalization arising from combinations of unit data that may not be "personal" in isolation or, importantly, fall under existing regulations.

The "Respect for Context" principle of the Consumer Privacy Bill of Rights touches on this, if interpreted dynamically. In its original formulation, the "context" is explained as the purposes and business processes that generate the data. In the big data world, those same processes also combine the data with other data – about the same individuals and about other individuals – in ways not always made transparent to each subject. This is a new kind of context, relevant to privacy and data policy in the same way other contexts are, but which may necessitate rewording or reformulating of the original principle text.

Mozilla's previous filing with the Federal Trade Commission articulated an Internet user's social graph as an example of aggregations of individual elements of less personal data that in combination become more personal.[2] Extending that example into the big data world, a combination of millions of Internet users' social graphs further changes the analysis. And yet, that combination is precisely what is being done be a number of private sector, intelligence, and law enforcement actors. Certainly, the subject matter of such work must be considered "personal" to some degree.

Ordinary web browsing activity includes a great deal of "personal" data, according to our users. For example, Firefox and other browsers store the list of URLs representing a

[1] Comments of Mozilla, Federal Trade Commission, Protecting Consumer Privacy (Feb. 23, 2011), http://www.ftc.gov/sites/default/files/documents/public_comments/preliminary-ftc-staff-report-protecting-consumer-privacy-era-rapid-change-proposed-framework/00480-58110.pdf
[2] *Id.* at 5.

user's web browsing history for that person's use and reference. An ordinary user views that information as "personal," though such metadata is outside the scope of regulation.

Over the past decade, very few new forms of PII have been articulated, but the scope of such "personal" information has grown. Additionally, machine-generated data built from ordinary activity, coupled with non-personal data, can uniquely identify a user and dynamically personalize her use of an online service, through pricing, advertising display, and other activities that feel very "personal."

Adapting and expanding the scope of "personal" data is an essential precursor to a proper understanding of privacy in a big data world.

B. Promoting Meaningful, Protected Portability

Responsive to Question 3: What technological trends or key technologies will affect the collection, storage, analysis and use of big data? Are there particularly promising technologies or new practices for safeguarding privacy while enabling effective uses of big data?

One trend is a growing number of silos, vertically and horizontally integrated services, applications, and devices that generate a diverse and broad amount of user data, and share and analyze that data across integrated and partner services. This is responsive to both question 3 and question 1, in that it is a significant technology development with ramifications for big data policy generally.

Integration has created a world where the concept of "big data" is salient even within a single company. Such integration can obviate the financial or technical need for portability of data, and make it easier, or cheaper, to lock data within a single organization. The result is often reduced transparency and user control over personal data and combinations of data with personal impact. It can also create policy problems above and beyond the traditional locus of privacy, transparency, and control issues, such as competition, innovation, and economic growth challenges.

The contrary vision is one in which data sets are portable and not locked within a single company or ecosystem. Portability necessitates a degree of transparency and enables control through choice of platform and environment.

Mozilla is committed to developing, promoting, and promulgating interoperable and standards-driven technologies, and to opposing silos and walled gardens of data and

technology. Further OSTP study of data portability and its relation to big data and to data and service silos would greatly help advance the vision of the Consumer Privacy Bill of Rights in the big data world.

C. Resisting Barriers and Balkanization

Responsive to Question 5: What issues are raised by the use of big data across jurisdictions, such as the adequacy of current international laws, regulations, or norms?

One opposite value to portability is integrated, locked ecosystems, as described above – pressures endogenous to the dataset and those processes operating on it. A related, yet highly distinct force is balkanization and division at geographic or regional boundaries – pressures generally exogenous from the data set itself. These forces have been increasing in past months and years as a result of significant economic opportunities for nations through internalizing larger segments of the Internet ecosystem, as well as defensive responses to revelations of expansive surveillance. These issues are responsive to question 5 in that they color global policy design and harmonization for big data, and create proximate challenges to building and using global big data systems.

External drivers that would mandate national borders for data or introduce restrictions to data portability represent a major concern for nascent big data policy and for current systems. As the OSTP process will no doubt reiterate time and time again, big data is at an early stage of understanding. Technical and legal mechanisms that hamstring its growth by imposing artificial and unnecessary barriers serving parochial interests must be avoided and resisted. OSTP can, and should, support further study of the harmful impact of such barriers, and the benefits of safe harbors and other mechanisms to advance open information flow across jurisdictions, and should advocate for open data flow within Administration policy processes and discussions.

D. Restoring Trust by Fixing Surveillance Practices

Responsive to Question 4: How should the policy frameworks or regulations for handling big data differ between the government and the private sector? Please be specific as to the type of entity and type of use (e.g., law enforcement, government services, commercial, academic research, etc.).

Government use of big data carries with it one major contextual difference from private sector use: there is usually no inherent optionality. Although caveats and qualifiers

abound, in general, people perceive a choice to use or not use services offered by the private sector, and therefore a fundamental ability to escape those data collection and use mechanisms. People do not have a choice to opt out of being subject to law enforcement or intelligence activities. This doesn't mean private sector use of big data doesn't raise major and legitimate concerns. But it does mean that private sector use is fundamentally different from government use. It means that some concerns arising from misuse of personal data in all its forms, as well as the harm of not knowing how data is being used, are heightened when it is the government using it. This is responsive to question 4 in that respecting these heightened concerns demands that government use, particularly those uses that are not optional, be viewed differently and separately from private sector use of big data.

Two issues have arisen in the context of government surveillance practices that are salient to this point. First, one of the major objectives behind ECPA reform efforts[3] is addressing the third party doctrine, the notion that data voluntarily ceded to a private sector company loses privacy rights with respect to future sharing of that data with the government. In a world where government use of data and private use of data present different normative balances of interests, this concept is out of date and needs to be changed - particularly where increasingly frequent National Security Letters (NSLs) result in divulgence of the privately held data without any legal pathway to inform the subject. Second, specific government surveillance conducted through interceptions of data center communications[4] represent a direct method for government access to private sector held data, without the knowledge or assistance of the company that collected and transferred the data. These issues render it difficult if not impossible to implement appropriate differences in policy between government and private sector access to, and use of, big data – counter to the policy need to respect heightened concerns associated with government use.

Overall, to the extent that the objective of this OSTP process is to encourage government and private sector collaborative efforts to shape policies and best practices for big data, and that establishing and defending trust in that ecosystem must be a key goal, surveillance and surveillance reform have a proximate, significant impact.

---

[3] *See, e.g.*, Rainey Reitman, "Deep Dive: Updating the Electronic Communications Privacy Act," *Electronic Frontier Foundation* (Dec. 6, 2012), at https://www.eff.org/deeplinks/ 2012/12/deep-dive-updating-electronic-communications-privacy-act.

[4] Barton Gellman and Ashkan Soltani, "NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say," *Washington Post* (Oct. 30, 2013), at http://www.washingtonpost.com/world/national-security/nsa-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story.html.

III. Moving Forward

The outcome of this OSTP process is unlikely to be any line drawing as to what is or is not a good practice with regards to big data – and that's proper at this stage, because much more work needs to be done to understand the context. Part of the path forward, therefore, is more research on technologies and policies of big data. Another part is continued multi-stakeholder engagement as the industry evolves, to avoid determining policies in a vacuum. OSTP has a facilitating, sponsoring, and convening role to play for both of these directions.

Collective understanding and development of big data policy is at an early stage, as are the technologies, options, and social, political, and market structures around big data. But it's not too early to have positive impact on the future of big data. Research and experimentation will help – on better architectures to advance the principles of transparency and control, on a better understanding of social perspectives around big data, and on legal and policy systems to improve trust. Changes in data policy generally – particularly around the evolving concept of "personal" data, surveillance practices, and growing challenges with portability and barriers – will have an impact wherever little data combines into big.

OSTP and every commenter participating in this proceeding have a mutual opportunity and a collective responsibility to work towards improving public awareness and literacy around big data, its technologies and policies. Meaningful, not merely superficial, public engagement with these issues as they develop would prove hugely helpful to advancing the public interest.

Finally, the big challenges of big data demand big thinking. Even in those dimensions that have yet to bear fruit, such as tagging data to improve transparency and control, it is too early to give up, and more investment may yet produce huge positive returns. Any combination of policy and law that can help make big data more tractable would be worth the efforts involved.