

April 4, 2014
Office of Science and Technology Policy
Eisenhower Executive Office Building
1650 Pennsylvania Ave. NW
Washington, DC 20502

Re: Big Data RFI—OTI comments on the White House’s Big Data Initiative

In response to the Office of Science and Technology Policy’s Request for Information on the implications of “big data,” the Open Technology Institute (OTI) at New America submits the following short essay, “The Biggest Data of All.” This essay was first delivered as a talk at New York University by OTI Policy Director Kevin S. Bankston as part of the workshop on “Social, Cultural & Ethical Dimensions of ‘Big Data,’” the second of three workshops co-hosted by the Office of Science and Technology Policy as a part of the White House’s big data initiative.

The Biggest Data of All

When considering the future of big data and privacy, we must consider the biggest data of all, the data set that encompasses almost all of the others: the data that transits the Internet.

As our offline activities and records move online—our shopping, our consumption of news and entertainment, our financial and legal and medical records and transactions, and an ever-increasing number of personal and business communications of every kind, even the most sensitive—the depth and breadth of this massive data set continues to expand. As all roads once led to Rome, today, nearly all data streams eventually flow into and through the great river of data that is the Internet.

Therefore, when considering the ethics of big data and privacy, it is necessary to look to the ISPs and governments—including our own—that have access to that river of data, often subject to unclear or insufficient legal restrictions.

What are the duties that these stewards of cyberspace owe to us? Would it be ethical, for example, for an ISP to secretly monitor and record everything you are reading and searching for online, so it could better serve you targeted ads or sell that information to data brokers? What if it gave you notice? Or even a discount on your service in exchange for permission?

And what of the government? Would it be right for our government to secretly install massive automated surveillance stations on top of major Internet exchange points inside the U.S., vacuuming up all of the data under questionable legal authorities, and filtering for suspicious identifiers and patterns? Would it be right for our government to secretly tap into the fiber lines that link the data centers of U.S. companies whose services are used by masses of innocents both here and abroad?

These questions are not hypothetical. And the ultimate report of the big data working group will be incomplete if it does not at least attempt to address some of these questions. There are three answers to these questions, in particular, that I hope we'll eventually see.

The first answer is transparency. Whether talking about ISPs or governments, tapping of the Internet backbone shouldn't occur without the knowledge and consent of us, the customers and the governed. That is why I hope that the Administration will soon finally admit to a fact that's been on the front pages of every major newspaper since December 2005, that's been evidenced by whistleblower documents leaked from inside of AT&T and the government, that's implicit but obvious in FISA court opinions and procedures that are now declassified, and that's even been admitted to by Senate Intelligence Committee Chair Dianne Feinstein.

For us to have a meaningful public debate in the public square, Congress, or the courts, the Administration must declassify the open secret that everyone already knows: the NSA is tapping the Internet backbone inside of the United States.

The second answer is encryption. Much of the big data discussion has focused on the risks to data in storage, and on the anonymization and

encryption tech that might protect that data. But we also must focus on encryption for data in transit—be it encryption that protects data sent between me and you, between me and your web site, between our email provider’s servers, or between Google and Yahoo’s data centers. When it comes to protecting our digital privacy, code is law, and encryption is one of the strongest laws on the books so long as we use it. Therefore, I urge this working group to conclude, as the President’s NSA Review Group did, that the U.S. government should strongly support rather than undermine the widespread use of encryption technology—not only for data at rest but also in transit.

A third and final answer to the problem of privacy when it comes to the biggest data of all is the much-needed re-evaluation of the distinction between communications content and non-content (or addressing or “meta”) data about those communications, where content has long been considered the most sensitive, such that non-content has been accorded little legal protection. As Danny Weitzner, the convener of the first in this series of workshops, testified to the Privacy & Civil Liberties Oversight Board: “Metadata at scale is at least as revealing as content.” Particularly in the Internet context, metadata can provide an intensely revealing portrait of one’s private life, including revealing facts or patterns of behavior that would never be revealed in the content of a communication and in many cases would not even be recognized by the person doing the communicating.

Countless legal and technical experts, including Justice Sotomayor, the oversight board, and the review group have called into question the continued validity of this distinction between content and metadata, and the review group specifically recommended that the government commission a study interrogating that distinction.

I hope that this working group will be the first step in such a study. I hope that this working group will highlight the importance of an encrypted Internet to the future of privacy and security. And I hope that this working group will be a force for greater transparency around how our data is collected and used, whether by our ISPs or by our National Security Agency.

Thank you,

Kevin S. Bankston
Policy Director, Open Technology Institute
New America
1899 L St NW, Suite 400
Washington, DC 20036