Big Data Before and After Snowden

Stephen I. Vladeck*

> "[M]y problem with paper is that all communication dies with it. It holds no possibility of continuity."[1]

What a difference a year makes.

On February 27, 2013, I had the privilege of moderating the closing panel at the *Journal of National Security Law and Policy*'s first annual symposium, a daylong event titled "Swimming in the Ocean of Big Data: National Security in an Age of Unlimited Information." Our panel was tasked with "Charting the Future: What to Expect from Big Data," and included Mary Ellen Callahan from Jenner and Block; Elisebeth Cook from WilmerHale (and the Privacy and Civil Liberties Oversight Board); John Grant from Palintir Technologies; Adam Isles from the Chertoff Group LLC; Greg Nojeim from the Center for Democracy and Technology; Robert O'Harrow from the *Washington Post*; and Marc Rotenberg from the Electronic Privacy Information Center.

Although this issue of the *Journal* includes an array of papers prepared in conjunction with the symposium, we decided on a somewhat different tack for the closing session. Given both the nature of the panel – which was set up more as a roundtable policy discussion than a series of seriatim paper presentations – and the diverse (and non-academic) backgrounds of our panelists, we chose to devote this portion of the symposium's print edition to a transcript of the wide-ranging discussion, edited only to omit various introductory and concluding remarks; to clean up speech disfluencies; and to add citations where appropriate. That transcript follows this Introduction, in which I seek to place our colloquy in context.[2]

In returning to the panel transcript as the one-year anniversary approaches, it is hard not to reflect and remark upon the fortuitous timing of our conversation. Just one day earlier, the U.S. Supreme Court had ruled 5-4 in *Clapper* v. *Amnesty International*[3] that an array of lawyers, journalists, and civil liberties and human rights groups lacked standing to challenge the constitutionality of

---

* Professor of Law and Associate Dean for Scholarship, American University Washington College of Law. Thanks to Denise Bell and Laura Donohue for their tireless efforts in organizing the symposium (and inviting me to participate), to the panelists – Mary Ellen Callahan, Beth Cook, John Grant, Adam Isles, Greg Nojeim, Bob O'Harrow, and Marc Rotenberg – for being such good sports, and to Nadia Asancheyev, for everything.

1. Dave Eggers, The Circle 186 (2013).

2. *See* Transcript, 7 J. Nat'l. Security L. & Pol'y 341 (2014).

3. 133 S. Ct. 1138 (2013). For more on *Clapper* and its continuing significance after Snowden, see Stephen I. Vladeck, *Standing and Secret Surveillance*, 9 I/S: J.L. & Pol'y for Info. Soc'y (forthcoming 2014).

section 702 of the Foreign Intelligence Surveillance Act (FISA),[4] with Justice Alito's opinion for the Court turning on the extent to which the plaintiffs could not show that the surveillance they feared was "certainly impending,"[5] and that their claims were based instead on an entirely "speculative chain of possibilities."[6]

That same week, the Supreme Court had heard oral argument in *Maryland* v. *King*, which culminated in the Court's 5-4 decision on June 3, holding that a state law that authorized law enforcement officials to collect without consent – and store – DNA samples from all individuals arrested for "serious offenses" did not violate the Fourth Amendment.[7] As the transcript indicates, the potential uses (and abuses) of a national DNA database – and whether the Fourth Amendment has anything to say about such mass data collection and storage – was not far from the minds of the panelists when we convened on February 27.[8]

Most intriguingly, although we couldn't possibly have known at the time, were the coming disclosures from former NSA employee Edward Snowden. Indeed, just over three months after the conference, a series of news stories based upon revelations by Snowden brought into the public eye a veritable bevy of controversial U.S. surveillance enterprises, including the PRISM program under section 702,[9] and the bulk telephone metadata program under section 215 of the USA PATRIOT Act.[10] Here, in the flesh, was clear and incontrovertible evidence not just of the extent to which the government was already knee-deep into the collection of "Big Data," but also of the controversial nature of such a haystack-before-the-needle[11] approach to information gathering. Simply put, thanks to Snowden, the future that our panel was tasked with charting unfolded infinitely faster – and, arguably, quite a bit differently – than anyone could have predicted that day.

And yet, for as much as the Snowden revelations could not have been anticipated, variations on many – if not most – of the themes that have come to dominate the public discourse that those disclosures have precipitated in the ensuing weeks and months can be found in our two-hour discussion. In this

4. Pub. L. No. 95-511, 92 Stat. 1783 (1978) (codified at 50 U.S.C. §§ 1801 *et seq.*). Section 702 was added to FISA by the FISA Amendments Act of 2008 (FAA), Pub. L. No. 110-261, § 101, 122 Stat. 2436, 2438-2448 (codified at 50 U.S.C. § 1881a).

5. *Clapper*, 133 S. Ct. at 1147.

6. *Id*. at 1150.

7. 133 S. Ct. 1958 (2013).

8. *See* Transcript, *supra* note 1.

9. *See* Barton Gellman & Laura Poitras, *U.S., British Intelligence Mining Data from Nine U.S. Internet Companies in Broad Secret Program*, WASH. POST, June 6, 2013, at A1.

10. Uniting and Strengthening America By Providing Appropriate Tools Required To Intercept and Obstruct Terrorism (USA PATRIOT) Act of 2001, Pub. L. No. 107-56, § 215, 115 Stat. 272, 287 (codified as amended at 50 U.S.C. § 1861).

11. One of the government's principal descriptive justifications for the metadata program is the need to collect the haystack in order to find the needle. *See, e.g.*, Rachel Levinson-Waldman, *The Double Danger of the NSA's "Collect It All" Policy on Surveillance*, GUARDIAN, Oct. 10, 2013, http://www.theguardian.com/commentisfree/2013/oct/10/double-danger-nsa-surveillance. The problem with this metaphor in the metadata program, however, is that it presupposes that a needle *exists*. It's something else entirely to collect the haystack just in case a needle might one day appear.

short Introduction, I offer brief reflections on three of them, before suggesting one place in which our reaction to Snowden may have missed – and could benefit from – the panel's larger point.

I. THE (COLLECTION) SHIP HAS SAILED

As the transcript indicates, most – if not all – of the panelists proceeded from the assumption that the ship has already sailed with regard to the *collection* of Big Data. That wasn't necessarily the consensus view when our panel convened; as some of the other papers in this very issue suggest, privacy and civil liberties advocates often argued that reform proposals should focus on curtailing the initial *interception* of data, whether by private parties or the government.[12] Indeed, at the time of the symposium, restrictions on how private parties or the government *used* such data once it was collected appeared to be no more than secondary measures, at least in comparison to more robust front-end *collection* restrictions.[13]

And yet, the public response to the Snowden disclosures has largely vindicated – for better or worse – the views of our panelists. Take the bulk telephone metadata program under section 215 of the USA PATRIOT Act as an example. Even those who have called for "ending" bulk collection by the government have not argued that such metadata should never be collected by *anyone*; rather, most of the more widely supported reform proposals would require that the data be stored by those entities who collected it (*e.g.*, telecommunications providers), or other non-governmental third parties, with the government only authorized to *access* the data upon a more specific, individualized showing of relevance. That *this* has been the focus of the reform conversation drives home the point: even after – and notwithstanding – Snowden, there does not appear to be any emerging consensus for dramatically restricting the types or volumes of data that is being collected on a daily basis. Instead, the crux of the debate has been devoted to recalibrating how that data can be used, and by whom.

Part of this reality may reflect Big Data's upside:

> big data evangelists insist that data-driven decisionmaking can now give us better predictions in areas ranging from college admissions to dating to hiring. And it might one day help us better conserve precious resources, track and cure lethal diseases, and make our lives vastly safer and more efficient. Big data is not just for corporations. Smartphones and wearable sensors enable believers in the "Quantified Self" to measure their lives in order to improve sleep, lose weight, and get fitter.[14]

12. *See, e.g.*, Joffe v. Google, Inc., 729 F.3d 1262 (9th Cir. 2013) (restricting Google's collection of openly accessible WiFi data).

13. For examples of this trend, consider the various essays published as part of the *Stanford Law Review Online*'s symposium on "Privacy and Big Data," especially Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41 (2013).

14. *Id*. at 41.

Indeed, in any number of ways, and despite the obvious and well-documented privacy implications, Big Data makes most of our lives easier (whether we'd like it to do so or not). As a result, we, as a society, may simply be unwilling to go back to the halcyonic past – to a time before apps that can, among other things, tell us the best (and worst) routes home based upon live and constantly updating traffic conditions, or provide us with coupons targeted to reflect both our and our neighbors' purchasing patterns at our local grocery store.

Indeed, the ubiquity of Big Data is entirely a function of its *utility*. Thus, as Big Data technologies continue to develop, and as Big Data accessibility increases, it should follow that Big Data will become only that much *more* useful to its producers and consumers – and, as such, that much harder to resist. And insofar as Big Data is increasingly useful, front-end constraints on the *generation* of Big Data will be increasingly difficult to justify, whether as a matter of internal corporate practice, industry norm, or state or federal regulation. Simply put, the data is going to be out there; as the Snowden disclosures have illuminated, the question will instead turn to how it may be utilized – and by whom.

II. THE UNDERAPPRECIATED PRIMARY/SECONDARY USE DISTINCTION

Use restrictions can come in all shapes and sizes. Who can access the data? In what circumstances can the data be accessed? What procedures are in place to ensure that those circumstances have in fact arisen? What mechanisms exist to punish those who violate these rules? All of these are important questions – and issues that have been at the forefront of the myriad FISA reform proposals currently working their way through Congress. But our panel seized upon another piece of the puzzle: Whatever use restrictions the underlying regulatory regime imposes, should there also be robust restrictions on "secondary" use of the data once it is accessed, in addition to the "primary" use. Put another way, if the government is not allowed to access the bulk telephone records it has collected under section 215 of the USA PATRIOT Act until it has "reasonable articulable suspicion" that a specific phone number is directly relevant to an ongoing terrorism investigation, should there be additional legal constraints in place to govern what happens once it *has* validly accessed that data, both initially and downstream?

Although our panelists would likely draw the lines in somewhat different places, at least some consensus emerged from our discussion with respect to secondary use restrictions: *First*, there *should* be substantive limits on secondary uses, although those limits will necessarily vary on a context-specific basis. Thus, for example, the Supreme Court's 5-4 decision in *Maryland* v. *King* on the permissibility of collecting DNA samples from all individuals arrested for "serious offenses" stressed the significance of the limited purposes for which

such DNA information could be used.[15] As Justice Kennedy explained for the majority,

> It is undisputed that law enforcement officers analyze DNA for the sole purpose of generating a unique identifying number against which future samples may be matched . . . . If in the future police analyze samples to determine, for instance, an arrestee's predisposition for a particular disease or other hereditary factors not relevant to identity, that case would present additional privacy concerns not present here.[16]

In cases like *King*, then, a secondary use restriction in the DNA context would be to preclude government officials from using the DNA for any purpose other than that which justified the collection of the data in the first place – *i.e.*, linking suspects to crime scenes through forensic evidence.

Second, and related, is the question of how long data should be retained. Whereas physical limits on storage capacity had historically exerted at least some influence on data retention, technological advances, again, have made it far easier to store increasingly larger quantities of data for longer periods of time. Thus, whether there comes a point after which data must be destroyed – to protect privacy and reduce the possibility of unauthorized use – should also become a larger focal point of conversations for reform. Indeed, as several of the panelists hinted, there is an important – if subtle – distinction between mandatory *retention* periods (*e.g.*, requiring telecom providers to preserve metadata for three years), and mandatory *destruction* requirements. Even though we might assume that data can and should be destroyed at the end of a mandatory retention period, it may nevertheless behoove policymakers to hardwire into any regulatory regime clear requirements for data destruction – and meaningful penalties for noncompliance – especially if and when there is a clear point beyond which the data is no longer of value for the primary use for which it was collected.

III. THE PUBLIC/PRIVATE DIVIDE IS ELUSIVE, BUT NOT ILLUSORY

Finally, the elephant in the room in any conversation about Big Data reared its head at various points throughout our panel: Can commentators and policymakers meaningfully distinguish between the legal and policy concerns that arise from the collection and use of Big Data by the private sector, as compared to those same concerns with regard to such collection and use by the government? Indeed, a common refrain of those less critical of the various NSA surveillance programs that were disclosed by Edward Snowden is the extent to which we should be far *less* troubled by government Big Data than we should be by Big Data in the hands of the private sector, since the government is

15. Maryland v. King, 133 S. Ct. 1958.
16. *Id*. at 1979 (citation omitted).

accountable in any number of ways in which the private sector is not. At a minimum, these commentators have argued that there is no meaningful distinction from a privacy perspective as between the two. Although the takeaway from the panel discussion was that such a distinction is certainly elusive, we also appeared to agree that it is not necessarily illusory.

To be sure, the government, unlike the private sector, is circumscribed in its collection and use of Big Data by the Constitution – and the Fourth Amendment, in particular. Yes, the Supreme Court has held that individuals do not have an expectation of privacy (and are therefore not entitled to Fourth Amendment protections) with respect to metadata that they voluntarily provide to third parties.[17] But that decision (which is older than I am) largely turned on the view that, in such cases, the government is merely privy to the same information as the phone company from which the data was obtained.[18]

Even on those terms, five Justices have recently questioned the continuing vitality of such an approach to privacy with respect to third parties,[19] and one district judge has specifically refused to follow such a precedent in the context of the bulk telephone metadata program.[20] But perhaps more significantly, the government's ability to aggregate across data streams today to create a mosaic of individuals' data puts it in a unique position vis-à-vis any single private sector firm, which would presumably only have access to its own internally generated data. That is to say, even if we have no expectation of privacy in information we voluntarily provide to third parties, there may be some expectation that those third parties do not turn around and commingle all existing data streams in ways that the government may have the technological capacity and wherewithal – if not the present legal authority – to attempt. As the panel discussion underscored, this is a potentially momentous distinction between the private sector and the government, albeit one that may still be largely theoretical.

Ultimately, whether or not the government's ability to aggregate across data streams ends up providing a basis upon which to revisit Fourth Amendment doctrine, it returns us to the theme of more vigorous use and retention restrictions – as a matter of statute, if not constitutional imperative. And, in the context of the private sector, it would also underscore the need for additional constraints on data *sharing*, at least with other private parties.

In the final analysis, Big Data may raise comparable concerns with regard to how it is collected and used by private firms as compared to the government, but the solutions will necessarily vary in direct proportion to the use and retention restrictions upon which we ultimately coalesce. And although those

17. *See, e.g.*, Smith v. Maryland, 442 U.S. 735 (1979).

18. *See id.* at 741-745.

19. *See* United States v. Jones, 132 S. Ct. 945, 957 (2012) (Sotomayor, J., concurring); *id.* at 961-964 (Alito, J., concurring in the judgment).

20. *See* Klayman v. Obama, 957 F. Supp. 2d 1 (D.D.C. 2013). *But see* ACLU v. Clapper, 959 F. Supp. 2d 724 (S.D.N.Y. 2013) (upholding the metadata program against a Fourth Amendment challenge).

restrictions will necessarily be case-specific, it should not be too difficult a proposition to conclude that the range of permissible uses will necessarily – and materially – differ as between the corporate world and Big Brother.

CONCLUSION

In retrospect, the topic of the *Journal*'s first annual symposium may seem unusually prophetic. If nothing else, the Snowden disclosures had the effect of bringing into mainstream public consciousness the very discussions about Big Data that academic and technological experts had been trying to have for the better part of the previous decade – and driving home exactly what kind of surveillance capabilities Big Data can facilitate.

But while the increasingly immaterial debate over the propriety of Snowden's leaks continues, the reform conversation has – disquietingly – focused on changes to the specific government surveillance programs Snowden's leaks revealed. From the robust reforms proposed by Senator Leahy and Congressman Sensenbrenner to the recommendations made by the Privacy and Civil Liberties Oversight Board and the President's own Review Group on Intelligence and Communications Technologies, most of the attention has been on individual surveillance authorities, and how they can better be curtailed and/or overseen. I've written elsewhere that, in the process, we've missed the most important lesson from the Snowden disclosures – *i.e.*, the extent to which the compromise solution Congress reached in the 1970s with regard to foreign intelligence surveillance has largely broken down.[21]

No less important is the larger lesson Snowden's leaks have to offer with regard to the potential and the pitfalls of Big Data. Whatever comes of the bulk telephone metadata program, it's only a matter of time before the government finds *other* streams of Big Data to mine under the guise of counterterrorism and national security – or, more cynically, before we find out about such endeavors. As the panel discussion that follows illuminates, what we've learned from Snowden may well just be the tip of the iceberg – underscoring the need to spend less time worrying about Snowden, and more time searching for a far more structural understanding of Big Data, and how, insofar as it is here to stay, policymakers can better circumscribe its vices.

21. *See* Steve Vladeck, *Does "Espionage Porn" Make Us Stronger?*, JUST SECURITY (Jan. 23, 2014), http://justsecurity.org/2014/01/23/espionage-porn-stronger/.

***