# A NEW HARM MATRIX FOR CYBERSECURITY SURVEILLANCE

OMER TENE*



## INTRODUCTION

In its 2009 Cyberspace Policy Review, the White House stated that "cybersecurity risks pose some of the most serious economic and

* Associate Professor, College of Management Haim Striks School of Law, Israel; Vice President of Research and Education, International Association of Privacy Professionals; Senior Fellow, Future of Privacy Forum. I would like to thank Kelsey Finch for expert research assistance and Sam Pfeifle for useful comments.

national security challenges of the 21st Century."[1] Then-Secretary of Defense Leon Panetta cautioned against a "cyber Pearl Harbor; an attack that would cause physical destruction and the loss of life . . . paralyze and shock the nation and create a new, profound sense of vulnerability."[2] Highlighting the cybersecurity risks introduced by the very communication networks that have boosted the economy, productivity and growth, former Federal Bureau of Investigation (FBI) Director Robert Mueller pointed out that invaders had overrun the Roman Empire by using the same roads that had been built to spread Roman civilization and influence. The Director of the National Security Agency (NSA) has recently expressed concern that legislative reaction to the Snowden revelations would undermine the nation's cybersecurity defenses.[3]

To respond to these perils, governments all over the world have been developing comprehensive programs and systems to boost cyber defenses. By monitoring or imposing requirements to monitor communications data, such programs and systems inexorably affect individuals' privacy. This presents policymakers with a formidable challenge: balancing cyber and national security risks against privacy and civil liberties concerns. This delicate balancing act must be performed against a backdrop of laws that are grounded in an obsolescent technological reality. Legal distinctions between communications content and metadata; interception and access to stored information; and foreign intelligence and domestic law enforcement – do not necessarily reflect the existing state of play of the Internet, where metadata may be more revealing than content, storage more harmful than interception, and foreign and domestic intelligence inseparable.

This essay proposes new parameters for analysis of the privacy impact of communications monitoring programs. It focuses on the cyber defense arena, although some of its observations may be useful in the foreign intelligence and domestic law enforcement spheres as well. Part One sets forth the traditional distinctions underlying existing surveillance laws and policies. Part Two describes the sea change caused by technological developments to both cybersecurity risks and civil liberties protections. Part Three proposes a new matrix for assessing privacy harms of communication monitoring programs. It suggests that purely

1. THE WHITE HOUSE, CYBERSPACE POLICY REVIEW: ASSURING A TRUSTED AND RESILIENT INFORMATION AND COMMUNICATIONS INFRASTRUCTURE iii, May 2009, *available at* http://www.whitehouse.gov/assets/documents/Cyberspace_Policy_Review_final.pdf.

2. Sec'y of Def. Leon E. Panetta, Remarks on Cybersecurity to the Business Executives for National Security (Oct. 11, 2012).

3. David E. Sanger, *N.S.A. Director Says Snowden Leaks Hamper Efforts Against Cyberattacks*, N.Y. TIMES (Mar. 4, 2014), http://www.nytimes.com/2014/03/05/us/politics/spy-chief-says-leaks-hamper-protection-against-cyberattacks.html.

automated monitoring raises less privacy concerns than human observation. This, in turn, implies that the focal point for triggering legal protections should be the moment the system focuses on an individual suspect. It recommends a shift away from the traditional content-metadata dichotomy towards a framework that assesses privacy risk based on the purpose of monitoring. Different rules and procedures would apply to monitoring activities depending on if they involve collection of evidence, intelligence gathering or cybersecurity defense. Finally, it advocates implementation of privacy by design through organizational, technological and legal mechanisms, to ensure a proper balance is struck between national security concerns and protection of individual rights.

## I: TRADITIONAL DISTINCTIONS

Over the past few decades, the regulation of government surveillance has been grounded in a number of legal distinctions that serve as proxies for the measurement of privacy and civil liberties harms. Specifically, lawmakers and courts[4] consider the monitoring of *content* to be more invasive than the monitoring of *communications data* (now called metadata);[5] and the *interception* of communications more invasive than access to *stored* information.[6] In addition, the demilitarization of domestic law enforcement meant that a clear distinction was drawn between *foreign intelligence* and domestic *law enforcement*;[7] and that surveillance of *United States persons* was subject to more stringent controls than *foreign surveillance*[8]

### *A. Content Versus Metadata*

The distinction between content and non-content (metadata) pervades surveillance law at both the constitutional and statutory levels.[9] Whereas *Katz v. United States* held that individuals have a "reasonable expectation of privacy" in the contents of their conversations,[10] under the

4. This essay focuses on United States law, although similar distinctions apply under the laws of many Western nations.

5. *See generally* U.S. CONST. amend IV; 18 U.S.C. §§ 2510–2522 (2000).

6. *See generally* 18 U.S.C. §§ 2701–2712 (2002); 18 U.S.C. §§ 2510–2522; 50 U.S.C. § 1861 (2006); CHARLES DOYLE, CONG. RESEARCH SERV., RS22406, NATIONAL SECURITY LETTERS IN FOREIGN INTELLIGENCE INVESTIGATIONS: A GLIMPSE AT THE LEGAL BACKGROUND (2014), *available at* https://www.fas.org/sgp/crs/intel/RS22406.pdf.

7. *See generally* United States v. U.S. District Court, 407 U.S. 297 (1972) (*Keith*); 50 U.S.C. §1861.

8. *See generally* 50 U.S.C. §§ 1881a–1881b (2008).

9. *See* Omer Tene, *What Google Knows: Privacy and Internet Search Engines*, 2008 UTAH L. REV. 1434 (2008).

10. Katz v. United States, 389 U.S. 347 360 (1967) (Harlan, J. concurring).

third party doctrine, as iterated in *Smith v. Maryland*, telephone users lacked a reasonable expectation of privacy in their metadata.[11] Holding that "a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties," the Supreme Court asserted that pen registers, which capture the telephone numbers a user dialed, "do not acquire the contents of communications" and are therefore subject to warrantless search.[12] Courts went on to apply the third party doctrine in a wide variety of circumstances involving the disclosure of information to trusted third parties. For example, in *United States v. Forrester*, the court held that a government request to an Internet service provider (ISP) for IP addresses of websites visited, to/from information for e-mails, and volume sent to or from an account did not trigger the Fourth Amendment.[13] The Ninth Circuit held that "the Court in *Smith* and *Katz* drew a clear line between unprotected addressing information and protected content information that the government did not cross here."[14]

The distinction between content and non-content is determinative not only pursuant to the Fourth Amendment but also under the Electronic Communications Privacy Act (ECPA).[15] The ECPA's Wiretap Act places strict controls on the interception of the *contents* of a communication, requiring a showing of "probable cause [to believe] that an individual is committing, has committed, or is about to commit a particular offense [and that] normal investigative procedures have been tried and have failed or reasonably appear to be unlikely to succeed if tried or to be too dangerous."[16] Unlawful interception of the contents of a communication is subject to an evidentiary exclusionary rule,[17] as well as a penalty of up to five years imprisonment.[18] The ECPA's Pen Register Act sets a lower standard for law enforcement access to *metadata*.[19] A court will grant a government request if "the information likely to be obtained by such installation and use is relevant to an ongoing criminal investigation."[20] It does not apply an exclusionary rule, and imposes a lesser penalty upon violations including imprisonment for up to one year.[21]

11. Smith v. Maryland, 442 U.S. 735, 743–45 (1979).
12. *Id.* at 741.
13. United States v. Forrester, 512 F.3d 500 (9th Cir. 2007).
14. *Id.* at 510.
15. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended at 18 U.S.C. §§ 2510–2520, 2701–2711, 3121–3127 (2000)).
16. *Id.* § 2518.
17. *Id.* § 2515.
18. *Id.* § 2511(4)(a).
19. *Id.* § 3127; § 3121.
20. *Id.* § 3123(a).
21. *Id.* § 3123(d).

### *B. Interception Versus Stored Communications*

Whereas probable cause warrants are required to authorize the interception of the content of a communication,[22] the ECPA's Stored Communications Act (SCA) applies a looser standard to sanction government access to stored communications.[23] There is no requirement for probable cause; it is enough to show "specific and articulable facts showing that there are reasonable grounds" to believe communications are "relevant" to a criminal investigation.[24] The SCA is a highly complex statute, setting forth different rules and procedures for government access depending on whether access is: voluntary or compelled; to data held by an "electronic communication service" or a "remote computing service"; to data held by an entity offering services "to the public" or not; to the contents of a communication; or to communication data.[25] Complexity aside, it is clear that the SCA "is much less protective than the Wiretap Act."[26] Like the Pen Register Act, the SCA lacks an exclusionary rule and imposes a smaller penalty for violation: up to one-year imprisonment.[27]

### *C. Foreign Intelligence Versus Domestic Law Enforcement*

In 1972, in what has become know as the *Keith* case, the Supreme Court recognized that "criminal surveillances and those involving domestic security" are distinct, and that "[d]ifferent standards may be compatible with the Fourth Amendment if they are reasonable both in relation to the legitimate need of Government for intelligence information and the protected rights of our citizens."[28] In *Clapper v. Amnesty International*, the Supreme Court explained, that "[a]lthough the *Keith* opinion expressly disclaimed any ruling 'on the scope of the President's surveillance power with respect to the activities of foreign powers,' it implicitly suggested that a special framework for foreign intelligence surveillance might be constitutionally permissible."[29] Similarly, the Church Committee findings that "[t]he Government has often undertaken the secret surveillance of citizens on the basis of their political beliefs, even when those beliefs posed no threat of violence or

22. *Id.* § 2518.
23. *Id.* § 2703(d).
24. *Id*.
25. *Id.* § 2702.
26. Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1701, 1723 (2004).
27. 18 U.S.C. § 2701(b) (2000).
28. *Keith*, 407 U.S. 297, 322–23 (1972).
29. Clapper v. Amnesty Int'l, 133 S. Ct. 1138, 1143 (2013) (quoting *Keith*).

illegal acts"[30] has led to the 1978 enactment of the Foreign Intelligence Surveillance Act of 1978 (FISA).[31]

FISA restricted the government's authority to use electronic surveillance inside the United States to obtain foreign intelligence from "foreign powers," and required the government to obtain a warrant or court order from a newly created Foreign Intelligence Surveillance Court (FISC) for certain foreign intelligence surveillance activities.[32] On the one hand, FISA attempted to introduce safeguards against the kinds of abuses that had been documented by the Church Committee; on the other hand, it sought to preserve the government's ability to protect the nation against external threats.[33] The Report of the President's Review Group on Intelligence and Communications Technologies, which was established in August 2013 as a response to the Snowden revelations (the Review Group),[34] explained that "[t]hese features of the system established by FISA reflect Congress' understanding at the time of the central differences between electronic surveillance for foreign intelligence purposes and electronic surveillance for traditional criminal investigation purposes."[35] Criminal enforcement is subject to close scrutiny emanating from Fourth Amendment protection against unreasonable search and seizure.[36] Foreign intelligence is largely liberated from the same restrictions, as it is not geared to produce

30. FINAL REPORT OF THE UNITED STATES SENATE SELECT COMMITTEE TO STUDY GOVERNMENTAL OPERATIONS WITH RESPECT TO INTELLIGENCE ACTIVITIES. S. REP. NO. 755, 94th Cong., 2d Sess., at 5 (April 29, 1976).

31. Foreign Intelligence Surveillance Act of 1978, codified as amended at 50 U.S.C. §§ 1801 to 1885c (1978).

32. FISA did not deal with the government's authority to engage in foreign intelligence activities *outside* the United States. To that end, President Ronald Reagan issued Executive Order 12333 in 1981, which specifies the circumstances in which the intelligence agencies can engage in foreign intelligence surveillance outside the United States. Exec. Order. No. 12,333, 46 Fed. Reg. 59,941 (Dec. 4, 1981).

33. PRESIDENT'S REVIEW GROUP, LIBERTY AND SECURITY IN A CHANGING WORLD: REPORT AND RECOMMENDATIONS OF THE PRESIDENT'S REVIEW GROUP IN INTELLIGENCE AND COMMUNICATIONS TECHNOLOGIES 64 (2013) [hereinafter REVIEW GROUP]; *see also* PRIVACY AND CIVIL LIBERTIES OVERSIGHT BOARD, REPORT ON THE TELEPHONE RECORDS PROGRAM CONDUCTED UNDER SECTION 215 OF THE USA PATRIOT ACT AND ON THE OPERATIONS OF THE FOREIGN INTELLIGENCE SURVEILLANCE COURT, Jan. 23, 2014, *available at* https://www.fas.org/irp/offdocs/pclob-215.pdf (hereinafter PCLOB REPORT).

34. THE WHITE HOUSE, PRESIDENTIAL MEMORANDUM -- REVIEWING OUR GLOBAL SIGNALS INTELLIGENCE COLLECTION AND COMMUNICATIONS TECHNOLOGIES, August 12, 2013, http://www.whitehouse.gov/the-press-office/2013/08/12/presidential-memorandum-reviewing-our-global-signals-intelligence-collec.

35. *Id.* at 67.

36. JOHNNY H. KILLIAN ET AL., CONG. RESEARCH SERV., THE CONSTITUTION OF THE UNITED STATES OF AMERICA: ANALYSIS AND INTERPRETATION: ANALYSIS OF CASES DECIDED BY THE SUPREME COURT OF THE UNITED STATES TO JUNE 28, 2002, S. DOC. NO. 108-17, at 1287 (2004), *available at* http://www.gpo.gov/fdsys/pkg/CDOC-108sdoc17/pdf/CDOC-108sdoc17.pdf.

evidence in a United States court of law. [37]

### *D. United States Person Versus Foreign Surveillance*

The FISA Amendments Act of 2008 (FAA) adopted different rules for international communications depending on whether the target of the surveillance was a "United States person" or a non-United States person.[38] The FAA provides that if the government targets a *United States person*[39] who is outside the United States, it must satisfy traditional FISA requirements, including the issuance of a FISC warrant based on a finding that there is probable cause to believe that the United States person is an agent of a foreign power.[40] At the same time, section 702 of the FAA states that if the target of foreign intelligence surveillance is a *non-United States person* who is "reasonably believed to be located outside the United States," the government need not meet the FISA requirements even if the interception takes place inside the United States.[41]

FISA additionally requires intelligence agencies to "minimize" any private information collected about United States persons, deleting data that is irrelevant for intelligence purposes before providing it to others.[42] Minimization procedures are adopted by the Attorney General and reviewed by the FISC.[43] Specifically, according to the terms of FISC orders:

> Before any of the results from queries may be shared outside NSA (typically with the FBI), NSA must comply with minimization and dissemination requirements, and before NSA may share any results from queries that reveal information about a United States person, a high-level official must additionally determine that the information "is in fact related to counterterrorism information and that it is necessary to understand the counterterrorism information or assess its importance."[44]

Additional protections are afforded United States persons under Section 215 of the FISA, which provides that an investigation of a

37. *Id.* at 1342–43.

38. Foreign Intelligence Surveillance Act of 1978 Amendments Act of 2008, Pub. L. No. 110-261, 122 Stat. 2436.

39. 50 U.S.C. 1801(i) (2011).

40. *Id.* § 1881b(c)(1).

41. *Id.* § 1881a (titled "Procedures for targeting certain persons outside the United States other than United States persons").

42. REVIEW GROUP, *supra* note 33, at 62, 146.

43. 50 U.S.C. § 1801(h) (2011).

44. REVIEW GROUP, *supra* note 33, at 100.

United States person may not be "conducted solely on the basis of activities protected by the First Amendment to the Constitution."[45]

## II: CHANGES

Developments in information technologies and national security risks have dramatically shifted both sides of the risk of harm balance: to national security on the one hand and to individual freedoms on the other hand. These changes have complicated the debate unleashed by the Snowden revelations around the legality and ethics of mass surveillance in an age where data hoarding has become rampant not only by governments but also by big data businesses.[46] In particular, emerging cybersecurity threats may require increasingly comprehensive programs for scanning mass quantities of information; yet such programs strain existing constitutional and legal frameworks.

### *A. National Security Threats*

Over the past decade, a range of factors has transformed the nature of national security threats and amplified tensions between operational personnel and counsel applying laws and regulations. First, military threats have transcended national borders and become pervaded by non-state actors. Traditional national security threats, which were once the domain of Cold War superpowers, are now "privatized" through terrorist networks. Similarly, cybersecurity threats, once focused on cyber superpowers such as China and Russia, have now spilled over to private organizations such as Anonymous and even lone-wolf hackers who are sometimes co-opted by hostile states. Unlike nuclear proliferation, which requires a deployment of resources unavailable to non-state actors, very rudimentary tools are sufficient to unleash potentially devastating cyber attacks.

Second, defending national infrastructure is complicated by the interdependence of military and private networks. Indeed, the Internet itself emerged from the military domain, as have technologies, software and applications such as mobile phones and encryption tools. The vulnerability of such networks and connected infrastructure presents a menacing threat to the functioning of society. This includes risks not only to military assets and critical infrastructure but also to peripheral networks and even individual devices, which could be used as discreet

45. 50 U.S.C § 1861(a)(2)(B).

46. Jules Polonetsky & Omer Tene, *Privacy and Big Data: Making Ends Meet*, 66 STAN. L. REV. ONLINE 25 (2013); Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63 (2012).

gateways to stage focused attacks. Lord Cameron of Dillington famously coined the phrase that "no society stands more than nine meals away from total anarchy."[47] He predicted that merely three days of disruption to British supermarket supply chains would be enough to send law and order reeling, at grave risk to the life and security of ordinary citizens. Despite this, neither the United Kingdom nor any other country categorizes its supermarket chains as critical infrastructure.[48]

Third, the distinction between domestic and foreign communications is muddled. Indeed, the crisis in international relations resulting from the Snowden revelations, particularly between the European Union and the United States, is rooted in the fact that large amounts of foreign communications are routed through or stored in the United States.[49] At the same time, a large volume of domestic communications has cross-border elements, such as connections with foreign websites, services or counterparties.

### *B. Individual Liberty Protections*

Technological progress has upended not only the nature of national security risks but also the legal basis for individual liberty protections. To begin with, it is no longer clear that the fundamentals of surveillance law remain sound and relevant to modern technological realities. The line between *content* and *metadata* has become notoriously hard to delineate. Moreover, it is no longer clear that acquisition of communication content remains more "harmful" to privacy than the capture of metadata. The study of an individual's social network, contacts and location whereabouts, not to mention Google searches or browser history, may be more telling than the contents of the individual's communications. Specifically, in the cybersecurity context, the distinction between content and metadata is frequently irrelevant, since cyber risks may be embedded in both content and metadata. The

47. 13 Nov. 2007, Parl. Deb., H.L. (2007) 425 (U.K.), *available at* http://www.theyworkforyou.com/lords/?id=2007-11-13b.386.2 ("Think about it. At the end of the first day, you probably start pilfering food to feed your crying children. The second day you probably travel a very long way because you have heard about a food source, but by the time you get there, of course, you have to fight the thousands of others who have also heard about it. On the third day there will be rats, mayhem and maybe even murder.").

48. PETER SOMMER & IAN BROWN, OECD/IFP PROJECT ON FUTURE GLOBAL SHOCKS: REDUCING SYSTEMIC CYBERSECURITY RISK (2011), *available at* http://www.oecd.org/gov/risk/46889922.pdf.

49. 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), http://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html.

metaphorical Trojan horse dwells in all seven layers of the architecture of communications systems.[50]

Second, it is no longer clear that lawful *real-time interception* of communications presents a greater risk of harm to individuals than government *access to stored data*. In a world of big data, where every crumb of information is cataloged and recorded, perhaps forever, stored data can be far more revealing than transitory, ephemeral communications.[51] Real-time interception and communication analysis—alongside data retention limitations—would leave a smaller privacy footprint than mass collection and later access to stored data. In this vein, the Review Group ruminated whether:

> [T]echnical collection agencies could make use of artificial intelligence software that could be launched onto networks and would be able to determine in real time what precise information packets should be collected. Such smart software would be making the sorting decision online, as distinguished from the current situation in which vast amounts of data are swept up and the sorting is done after it has been copied on to data storages systems.[52]

Finally, the distinction between *foreign intelligence* and *domestic law enforcement* has blurred, as national security risks infiltrate domestic borders through the activities of rogue actors on the ground as well as the spread of malicious code to critical infrastructure. The Review Group pointed out that "[t]oday, no battlefield lines or Iron Curtain separates the communications in combat zones from the rest of the world. A vulnerability that can be exploited on the battlefield can also be exploited elsewhere."[53] It also noted that "[i]n recent decades, the global nature of the Internet has enabled daily cyberattacks on the communications of government, business, and ordinary Americans by hackers, organized crime, terrorists, and nation-states."[54] The Review Group stressed that the convergence of military and civilian systems for cybersecurity has profound implications, as information assurance for the military relies increasingly on the civilian sector, and the military and government rely on a broad range of critical infrastructure, which is mostly owned and operated by private parties.[55] This makes effective defense of private

50. *See* OSI Model, WIKIPEDIA, http://en.wikipedia.org/wiki/OSI_model (last visited Oct. 8, 2014).

51. Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW J. TECH. & IP 239 (2013).

52. REVIEW GROUP, *supra* note 33, at 174.

53. *Id.* at 187.

54. *Id.* at 184.

55. *Id.* at 180–87.

sector cyber infrastructure critical to military and other government functions.

## III: A NEW MATRIX

A response to these changes requires the development of a new matrix to assess risks to privacy and civil liberties that may result from government surveillance. This section argues that automated processing of bulk data should be viewed differently than individualized surveillance; that the distinction between content and metadata needs to be linked to the purpose of monitoring; and that surveillance—particularly on a mass scale—must be buttressed by measures of privacy by design.

### *A. Automated Processing*

The architecture of modern communications networks generally, and the Internet specifically, is based in monitoring, documentation and retention of data. Communications data is scanned by automated systems and retained for various purposes such as billing or defense from legal claims. Even the content layers of communications are regularly scanned for various purposes ranging from network security (*e.g.*, antivirus programs), user protection (*e.g.*, anti-spam or anti-phishing software), bandwidth management (*e.g.*, use of deep packet inspection by Internet service providers), or—more controversially—online behavioral advertising. Indeed, any digital data (except, perhaps, encrypted data) interacts with the systems processing it at some level.

Privacy concerns typically arise only when a system "zooms in" to focus on a particular individual. Two caveats are in order: first, to be sure, in an age of big data any mass scale surveillance operation is performed automatically. Gone are the days of switchboard operators listening in to calls. The proverbial Eve in data security parlance is always mechanical. Yet this does not imply that the distinction between automated and human monitoring is a red herring. The distinction remains meaningful so long as automatic monitoring does not single out an individual for special scrutiny or treatment. For example, the privacy effect of the London Underground's always-on CCTV cameras is minimal so long as it does not focus on the actions of any specific individual. Second, regardless of their impact on privacy, automated systems can certainly infringe *other* individual rights and interests, such as equality, fairness and due process. Indeed, the European privacy directive views automated decision-making as inherently suspect and grants every person the right "not to be subject to a decision which produces legal effects . . . which is based solely on automated

processing . . .".[56] Automated decisions can be arbitrary, unfounded, or reflect programmers' discrimination and bias. Consider, for example, an automated decision to lower an individual's credit rating based on inaccurate information or include her in a "no fly" list based on a mismatched name. Such decisions may very well be harmful, yet they do not necessarily infringe individuals' privacy.

In the private sector, automated scanning of communications data, including content, is common in various contexts. Google, for example, scans the email of more than 400 million Gmail users to tailor targeted ads to the contents of their correspondence.[57] Users (more or less) knowingly accept the benefit of the bargain, free storage space and a snappy user interface in return for use of their personal information for ad targeting.[58] While consent can be cited as a legal basis for Google's scanning of its users' correspondence, it cannot legitimize the monitoring of email of third parties who are not Gmail users themselves.[59] Apparently, the legitimacy of Google's actions is derived not only from users' consent but also from a restrictive view of the privacy implications of strictly automated monitoring.[60]

In its terms of service, Google states that the monitoring of Gmail content is "completely automated and involves no human review."[61] Similarly, the terms of service for Google Apps state that: "[i]t's important to note that our scanning and indexing procedures are 100% automated and involve no human interaction."[62] Such statements reflect

56. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281) 31, Art. 15. *See also* Omer Tene & Jules Polonetsky, *Judged by the Tin Man: Empowering Individuals in an Age of Big Data*, 11 J. TELECOMM. & HIGH TECH. L. 351 (2013).

57. Joel Rosenblatt, *Is Google Too Big to Sue Over Gmail Privacy Concerns*, BUSINESSWEEK (Mar. 6, 2014), http://www.businessweek.com/articles/2014-03-06/google-fights-gmail-privacy-lawsuits-bid-for-class-action-status.

58. Ashwin Seshagiri, *Claims that Google Violates Gmail User Privacy*, N.Y. TIMES (Oct. 1, 2013), http://www.nytimes.com/interactive/2013/10/02/technology/google-email-case.html.

59. *Id.*

60. *But see* Order Granting in Part and Denying in Part Defendant's Motion to Dismiss [Redacted], *In re* Google Inc. Gmail Litigation, Case No. 13-MD-02430-LHK (N.D. Ca. Sept. 26, 2013).

61. Defendant Google, Inc.'s Motion to Dismiss Plaintiffs' Consolidated Individual and Class Action Complaint; Memorandum of Points and Authorities in Support Thereof at 13, In re Google, Inc. Gmail Litigation, Case No. 13-MD-02430-LHK (N.D. Ca. Sept. 26, 2013).

62. Ashish Karve, *Google's Privacy Policy*, LAW FIRM MGMT CONSULTING (May 21, 2012), http://www.lawfirmmanagementconsulting.com/googles-privacy-policy. The official Google policy statement has since been revised. *Your Security and Privacy*, GOOGLE HELP CENTER, https://support.google.com/a/answer/60762?hl=en (last visited Mar. 17, 2014) (stating "Our systems scan and index emails and some other user data for multiple purposes; this scanning is 100% automated and cannot be turned off.").

an assumption that individuals are concerned less with automated monitoring than with human surveillance.

Similarly, employers regularly deploy data loss prevention (DLP) software to prevent security breaches or loss of trade secrets and intellectual property.[63] DLP systems scan communications *en masse*, raising red flags when they encounter suspicious activity, such as an email of a bank employee containing strings of 16 numbers, which could imply an attempt to leak credit card numbers.[64] Here too, an employee's consent provides a fragile legal basis for communications monitoring, given the inherent imbalance of power between employers and employees.[65] Something else must be at work here: a perception of content monitoring as proportional and legally justified so long as it does not extend to non-work related purposes.

In time, more and more systems and devices will monitor individuals' communications and behavior.[66] The Internet of Things connects to the network a broad array of devices and objects such as cars, thermometers, electric grids, garbage dispensers, and digital signage, for seamless and multilayered monitoring of individual data, including information about who we are and what we do.[67] The Internet of Things is infused with sensors and devices that continuously scan, collect, process and distribute data, requiring a reconceptualization of what it will mean for individuals to have privacy *vis-à-vis* such automated machines.[68] And while the privacy implications of the Internet of Things are just beginning to be understood and grappled with, it is clear that the level of automated processing of personal data is on the rise.[69]

63. Rich Mogull, *Understanding and Selecting a Data Loss Prevention Solution*, SANS INSTITUTE, https://securosis.com/assets/library/reports/DLP-Whitepaper.pdf (last visited Mar. 17, 2014).

64. *Id.*

65. *See, e.g.*, ARTICLE 29 DATA PROTECTION WORKING PARTY, OPINION 15/2011 ON THE DEFINITION OF CONSENT at 13, *available at* http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.pdf.

66. *See Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018*, CISCO (Feb. 5, 2014), http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html ("By the end of 2014, the number of mobile-connected devices will exceed the number of people on earth, and by 2018 there will be nearly 1.4 mobile devices per capita.").

67. *See* Marc Ferranti, *Beyond the Hype: Internet of Things Shows up Strong at Mobile World Congress*, PCWORLD (Feb. 27, 2014, 6:51 AM), http://www.pcworld.com/article/2102761/the-internet-of-things-beyond-the-hype-at-mobile-world-congress.html.

68. *See, e.g.*, Justin Brookman, Joseph Lorenzo Hall & G.S. Hans, Letter Re: Comments after November 2013 Workshop on the "Internet of Things" (Jan. 10, 2014), *available at* https://www.cdt.org/files/pdfs/iot-comments-cdt-2014.pdf.

69. *See, e.g.*, FED. TRADE COMM'N, INTERNET OF THINGS - PRIVACY AND SECURITY IN A CONNECTED WORLD, Nov. 2013, *available at* http://www.ftc.gov/news-events/events-calendar/2013/11/internet-things-privacy-security-connected-world.

The interaction between individuals and non-sentient devices is not restricted to the digital world. A weight scale can record individuals' weight and a mirror captures their image; yet few (healthy) individuals would regard such interactions as privacy infringing. As long as the weight or image is not transmitted to other humans and used to humiliate, discriminate or otherwise harm an individual, users' privacy is not affected.[70] The human-machine interaction, in and of itself, does not infringe on privacy any more than the gaze of a household cat. While scholars have struggled for many years to conceptualize privacy, most would agree that the notion involves a delineation of borders between an individual and society.[71] Privacy intrusions entail exposure to another individual or group of individuals, such as a community, society or nation, who may be curious, judging, discriminating, degrading or harmful toward the individual in various other ways.[72] Privacy protects individuals from other individuals—not from machines.

Matthew Tokson argues against expanding the notion of privacy to protection from automated processes.[73] Against the backdrop of the third party doctrine, he proposes to distinguish between information willfully disclosed to a third party *human* and information made available for *automated* processing.[74] Only the former type of disclosure should be regarded as a revocation of a "reasonable expectation of privacy," whereas exposure of information to a machine does not involve a voluntary assumption of privacy risk.[75] Hence, according to Tokson, the conflation of human and automated monitoring is a double-edged sword, which can be used to legitimize a broad swath of exemptions to the Fourth Amendment.[76] Recognizing that data crunching by a machine does not in itself infringe privacy provides a basis to argue that Fourth Amendment protection should expand to data automatically shared with ISPs, banks, email providers, cloud vendors and a plethora of other automated third parties, as an inevitable byproduct of life in a digital world.[77]

In this vein, Orin Kerr writes: "The best answer is that a search occurs when information from or about the data is exposed to possible

70. *See* ALAN WESTIN, PRIVACY AND FREEDOM (1967) (defining privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.").

71. *Id.*

72. *Id.*

73. *See* Matthew J. Tokson, *Automation and the Fourth Amendment*, 96 IOWA L. REV. 581 (2010).

74. *Id.* at 629.

75. *Id.* at 638.

76. *Id.* at 586.

77. *Id.* at 647.

human observation, such as when it appears on a screen, rather than when it is copied by the hard drive or processed by the computer."[78] Similarly, Tokson argues:

> [W]ithout some modicum of human observation, disclosure of our information to automated systems alone is ultimately no different from 'disclosure' to any other inanimate object that stores our personal data. Automated computers alone do not 'observe' us . . . These devices cannot see us, think about us, judge us, ridicule us, or be curious about us—they cannot perceive us at all. They cannot, then, truly violate our privacy.[79]

A case in point is the Transportation Security Administration's (TSA) decision in 2011 to eliminate the actual image of passengers projected to TSA operatives from body scanners at airports and to replace them with a generic outline of a person.[80] The TSA noted, "By eliminating the image of an actual passenger and replacing it with a generic outline of a person, passengers are able to view the same outline that the TSA officer sees. Further, a separate TSA officer will no longer be required to view the image in a remotely located viewing room."[81] Clearly, an automatic body scanner machine remains capable of "seeing" individual passengers naked, but absent the projection of such an image to a human TSA officer, the impact on individuals' privacy is limited.

A line of Supreme Court cases dealing with dog sniffs in airports and police stops supports a distinction between automated and human monitoring. In *United States v. Place*, the defendant challenged the legality of a police dog's sniff test, which detected the presence of narcotics in his luggage at an airport.[82] The Supreme Court held:

> [A] 'canine sniff' by a well-trained narcotics detection dog . . . does not require opening the luggage. It does not expose non-contraband items that otherwise would remain hidden from public view, as does, for example, an officer's rummaging through the contents of the luggage. . . Thus, despite the fact that the sniff tells the authorities something about the contents of the luggage, the information obtained is limited. This limited disclosure also ensures that the

78. Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531, 551 (2005).

79. Tokson, *supra* note 73.

80. Press Release, Transp. Sec. Admin., TSA Takes Next Steps to Further Enhance Passenger Privacy (July 20, 2011), http://www.tsa.gov/press/releases/2011/07/20/tsa-takes-next-steps-further-enhance-passenger-privacy.

81. *Id.*

82. *See* United States v. Place, 462 U.S. 696 (1983); *also see* Illinois v. Caballes, 543 U.S. 405 (2005).

> owner of the property is not subjected to the embarrassment and inconvenience entailed in less discriminate and more intrusive investigative methods.[83]

Hence, the Court held that the non-human monitoring of the content of the defendant's luggage did not constitute a Fourth Amendment search.[84] Indeed, the Court emphasized that "the manner in which information is obtained through this investigative technique is much less intrusive than a typical search."[85] Similar logic can apply to "packet sniffers,"[86] which, like police dogs in airports, monitor communications data for illicit, malicious activities.

A distinction between wholesale automated monitoring and individualized scrutiny can help adapt the *Smith v. Maryland* rationale to a reality of ubiquitous data collection. In the context of the Snowden revelations, commentators have argued that the government's reliance on *Smith*, with its trivial collection of phone numbers on one individual's pen register, to justify the NSA's compilation of the metadata of an entire nation is specious.[87] Yet this criticism ignores the fact that in *Smith*, law enforcement "zoomed in" to focus on one suspect; this should have triggered Fourth Amendment protection, since the privacy impact was pronounced. In comparison, wholesale collection of metadata without an individual focus leaves less of a footprint on any specific individual's privacy. So while *Smith*'s application of the third party doctrine may have been unwarranted, this does not necessarily discredit the government's arguments in favor of collecting bulk metadata based on FISA Section 215.[88]

The approach outlined here toward automated processing closely tracks the pervasive discussions among privacy policymakers around the concept of de-identification or anonymity. De-identification is rooted in a notion of privacy as an ability to hide in a crowd.[89] It reflects a belief that

83. *See Place*, 462 U.S. at 707.

84. *Id.*

85. *Id.*

86. *See* Adrian Hannah, *Packet Sniffing Basics*, LINUX JOURNAL (Nov. 14, 2011), http://www.linuxjournal.com/content/packet-sniffing-basics.

87. *See, e.g.*, Andrew Cohen, *Is The NSA's Spying Constitutional? It Depends Which Judge You Ask*, ATLANTIC (Dec. 27, 2013, 2:57 PM), http://www.theatlantic.com/national/archive/2013/12/is-the-nsas-spying-constitutional-it-depends-which-judge-you-ask/282672/; Jim Harper, *If You Think* Smith v. Maryland *Permits Mass Surveillance, You Haven't Read* Smith v. Maryland, CATO INSTITUTE (Aug. 20, 2013, 1:04 PM), http://www.cato.org/blog/you-think-smith-v-maryland-permits-mass-surveillance-you-havent-read-smith-v-maryland.

88. *See generally* David S. Kris, *On the Bulk Collection of Tangible Things*, LAWFARE RESEARCH PAPER SERIES (September 2013), http://www.lawfareblog.com/wp-content/uploads/2013/09/Lawfare-Research-Paper-Series-No.-4-2.pdf.

89. *See* Arvind Narayanan, *About 33 Bits*, 33 BITS OF ENTROPY,

individuals only lose privacy when they become the subject of attention. This is manifest in scientific techniques for masking the relationship between data and specific individuals, such as k-anonymity[90] and differential privacy.[91] With *k-anonymity*, individual attributes are suppressed or generalized until each row in a database is identical to at least k-1 other rows, at which point the database is said to be k-anonymous.[92] It allows individuals to blend into a crowd that is just large enough to prevent their re-identification.[93] *Differential privacy* emphasizes not whether an individual can be directly *associated* with a particular revealed value, but rather the extent to which any revealed value *depends* on an individual's data.[94] It avoids the ailments of de-identification by allowing data sharing in a way that maintains data quality while at the same time preserving individuals' privacy.[95] It enables organizations to share derivative data without subjecting any individual to more than a minimal risk of harm from the use of his or her data in computing the values to be released, even when those values are combined with other data that may be reasonably available.[96]

Robustly de-identified data is exempt from privacy regulation. According to the Federal Trade Commission (FTC), its privacy framework applies to "all commercial entities that collect or use consumer data that can be reasonably linked to a specific consumer, computer, or other device."[97] The FTC clarifies that data is not "reasonably linkable" to the extent that a company takes reasonable measures to ensure that it is de-identified; publicly commits not to try to re-identify the data; and contractually prohibits downstream recipients from trying to re-identify the data.[98] Similar limitations on the scope of

http://www.33bits.org/about (last visited Mar. 17, 2014).

90. Latanya Sweeney, *k-Anonymity: A Model for Protecting Privacy*, 10 INT'L J. ON UNCERTAINTY, FUZZINESS AND KNOWLEDGE-BASED SYSTEMS 557 (2002).

91. Cynthia Dwork, *Differential Privacy*, AUTOMATA, LANGUAGES & PROGRAMMING (2006).

92. *See* Sweeney, *supra* note 90.

93. *See id.*

94. *See* Dwork, *supra* note 91.

95. *See id.*

96. Ruth Gavison gives the following example, which helps illustrate the point: "Consider the famous anecdote about the priest who was asked, at a party, whether he had heard any exceptional stories during confessionals. 'In fact,' the priest replied, 'my first confessor is a good example, since he confessed to a murder.' A few minutes later, an elegant man joined the group, saw the priest, and greeted him warmly. When asked how he knew the priest, the man replied: 'Why, I had the honor of being his first confessor.'" Ruth Gavison, *Privacy And The Limits Of Law*, 89 YALE L. J. 421, 430-31 (1980).

97. FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS vii (Mar. 2012), *available at* http://ftc.gov/os/2012/03/120326privacyreport.pdf.

98. *Id.* at iv.

data regulation apply under a plethora of United States and European privacy laws.[99] For the purposes of this essay, the point is that just like with de-identified data, as long as no individual is targeted and subject to special scrutiny, automated processing implies limited privacy risk.

### *B. Content / Non-content*

The ECPA defines the "contents" of a communication as "any information concerning the substance, purport, or meaning of that communication."[100] Alas, the distinction between content and non-content (metadata) is increasingly fading, and it is not clear that content remains a good indicator of privacy harm.[101] It is easy enough to distinguish between the content of a letter and the non-content information disclosed on its envelope. Similarly, the content of a telephone call cannot be confused with metadata, such as the calling number, the number called, and the call's duration.[102] To be sure, telephone companies typically held additional information about their subscribers, such as their contact details, payment methods, and usage record.[103] Such data could be used to draw useful conclusions about a subscriber's demographics, age, social and professional network, and more.[104] While such information implicated a subscriber's right to privacy, it did not infiltrate the personal core of the subscriber's conversations like the content of his communication.[105] For example, a telephone company could have known that a subscriber called a certain other subscriber, who was not his spouse, several times daily. But it could not tell whether the parties discussed work issues, exchanged

99. *See generally* Paul M. Schwartz, & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L.Q. REV. 1814 (2011).

100. 18 U.S.C. § 2510(8) (2002).

101. *See* Jane Mayer, *What's the Matter with Metadata?*, NEW YORKER (June 6, 2013), http://www.newyorker.com/online/blogs/newsdesk/2013/06/verizon-nsa-metadata-surveillance-problem.html.

102. *See* Doug Aamoth, *Verizon, Telephony Metadata, the National Security Agency and You*, TIME TECH (June 6, 2013), http://techland.time.com/2013/06/06/verizon-telephony-metadata-the-national-security-agency-and-you/.

103. *See Verizon Transparency Report*, VERIZON (Jan. 22, 2014), http://transparency.verizon.com/themes/site_themes/transparency/Verizon-Transparency-Report-US.pdf.

104. *See* Anton Troianovski, *Phone Firms Sell Data on Customers*, WALL ST. J., (May 21, 2013, 7:13 PM), http://online.wsj.com/news/articles/SB10001424127887323463704578497153556847658.

105. Transcript of President Obama's Remarks on NSA Controversy, WASH. WIRE, June 7, 2013, http://blogs.wsj.com/washwire/2013/06/07/transcript-what-obama-said-on-nsa-controversy ("When it comes to telephone calls, nobody is listening to your telephone calls. That's not what this program's about. As was indicated, what the intelligence community is doing is looking at phone numbers and durations of calls. They are not looking at people's names, and they're not looking at content.").

recipes, or nurtured intimate relations.

The shift from analog to digital communications has eroded the binary nature of content and metadata. Consider email, a basic form of digital communications. Most would agree that the body of an email constitutes the content of a communication whereas information about an email's size (*e.g.*, two megabytes) or text length (*e.g.*, 12 lines) was metadata. But what about the subject line? While technically, the subject line belongs to the email's header and is therefore apparently metadata, it potentially (but not always) conveys the content of a communication.[106] And what about an algorithmically calculated "risk score" assigned to the contents of an email message based on parameters that could be purely content-based (*e.g.*, keywords), metadata-based (*e.g.*, the size of the message), or a combination of both (*e.g.*, number of words in a message, its language, or heuristic models based on its syntax).[107] And what about analysis of whether or not a message is encrypted? Apparently the mere fact that data is encrypted constitutes metadata; but making that determination requires analysis of information in the content layer.[108]

The analysis becomes even thornier with respect to individuals' online browsing habits. Courts have struggled to determine whether IP addresses,[109] URLs[110] or search query logs[111] should be treated as content or metadata. For several years, courts have wrestled with categorizing a list of URLs visited by an individual as content or metadata.[112] In a footnote to its *Forrester* decision, the Ninth Circuit expressed its concern about the ability of URLs to reveal content but concluded that IP addresses "reveal no more about the underlying contents of communications than do phone numbers" and are constitutionally indistinguishable from them.[113]

106. *In re* Application of the United States for an Order Authorizing the Use of a Pen Register, 396 F.Supp.2d 45, 48 (D. Mass. 2005).

107. *See, e.g.*, Method to calculate a risk score of a folder that has been scanned for confidential information, U.S. Patent No. 8,516,597 (filed Mar. 1, 2011), *available at* http://www.google.com/patents/US8516597.

108. *See, e.g.*, Andrei Robachevsky et al., *Position Paper Submission for W3C/IAB Workshop on Strengthening the Internet Against Pervasive Monitoring (STRINT)*, WORLD WIDE WEB CONSORTIUM at 3 (Dec. 2013), *available at* https://www.w3.org/2014/strint/papers/06.pdf.

109. United States v. Forrester, 495 F.3d 1041, 1048-49 (9th Cir. 2007) (IP addresses are metadata).

110. *Id.* at n. 6 (URLs "might be more Constitutionally problematic").

111. *In re* Application of the U.S. for an Order Authorizing the Use of a Pen Register, 396 F.Supp.2d 45, 49 (D. Mass. 2005) (search queries constitute content). *See also* Jayni Foley, *Are Google Searches Private? An Originalist Interpretation of the Fourth Amendment in Online Communication Cases*, 22 BERKELEY TECH. L.J. 447 (2007); Tene, *supra* note 9.

112. Matthew J. Tokson, The *Content/Envelope Distinction in Internet Law*, 50 WM. & MARY L. REV. 2105 (2009).

113. United States v. Forrester, 512 F.3d 500, 510 n.6 (9th Cir. 2007); *see also In re*

On the one hand, the data in a URL does not reflect "the substance, purport or meaning"[114] of a communication. On the other hand, it might give away the content, such as in the case of the URL <http://time.com/18691/edward-snowden-talks-privacy-and-security-at-sxsw-interactive>.[115] Commentators too have split over the distinction in this context, with some arguing for an expansion of privacy through a broad interpretation of "content,"[116] while others support a more restrictive approach.[117] Similarly, the most recent edition of the Department of Justice Manual on Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations struggles to delineate the content-noncontent divide.[118] For example, with respect to URLs, the Manual notes:

> In some circumstances, questions may arise regarding whether particular components of network communications contain content. . . Because of these and other issues, the United States Attorneys' Manual currently requires prior consultation with [the Computer Crime and Intellectual Property Section] before a pen/trap may be used to collect all or part of a URL.[119]

The forward march of technology has changed how content-like metadata is. This is evident particularly with online search queries, which have made it much easier to draw associations from things like IP addresses and URLs. Hence, in a previous article, I argued that users' online search-query logs should be classified as communication contents.[120]

Additional questions arise with respect to information collected by cellular networks, particularly with respect to subscribers' geolocation.

Application of the U.S. for an Order Authorizing the Use of a Pen Register, 396 F.Supp.2d 45, 48 (D. Mass. 2005).

114. 18 U.S.C. § 2510(8) (2002).

115. The answer may be different for the domain name <http://time.com>; although the following domain name is more telling: <http://singlemenlookingforlove>.

116. Solove, *supra* note 26.

117. *See, e.g.*, Christopher Slogobin, *Transaction Surveillance by the Government*, 75 MISS. L.J. 139, 153 (2005); Rich Haglund, *Applying Pen Register and Trap and Trace Devices to Internet Communications: As Technology Changes, Is Congress or the Supreme Court Best-Suited To Protect Fourth Amendment Expectations of Privacy?*, 5 VAND. J. ENT. L. & PRAC. 137, 141-42 (2003); Christian David Hammel Schultz, *Unrestricted Federal Agent: "Carnivore" and the Need To Revise the Pen Register Statute*, 76 NOTRE DAME L. REV. 1215, 1241–42 (2001).

118. OFFICE OF LEGAL EDUC. EXEC. OFFICE FOR UNITED STATES ATT'YS, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS 151-53 (2009), *available at* http://www.justice.gov/criminal/cybercrime/docs/ssmanual2009.pdf.

119. *Id.* at 153.

120. Tene, *supra* note 9, at 1482.

On the one hand, such information is not part of "the substance, purport or meaning"[121] of a communication. On the other hand, geolocation information is created not only by automatic multilateration or GPS signal but also through more content-based methods such as check-ins on social networks or searches on Google Maps.[122]

Not only has the task of defining content and metadata become daunting, but it also lost some of its normative appeal. It is no longer clear that content is a robust indicator of privacy harm. On the contrary, communication contents may be no more private, sensitive or revealing than metadata.[123] Indeed, many experts claim that the richness and depth of metadata could be *more* informative than communication content.[124] In their book on wiretapping, Diffie and Landau state that "traffic analysis, not cryptanalysis, is the backbone of communications intelligence."[125] And if this is true for ordinary individuals, it is certainly true for terrorists or members of criminal organizations who no doubt *assume* that their communications are being analyzed. Such rogue players can use various methods, such as encryption or steganography, to conceal the purport and meaning of their communications even in plain sight.[126] However, it is far more difficult for them to hide their location, social network, or online usage habits, and the very act of concealing such information could raise suspicion.[127]

121. 18 U.S.C. § 2510(8) (2002)

122. *See* JD Lasica, *Beyond Foursquare: Geolocation Services Proliferate, Mature*, PBS.ORG (Feb. 28, 2013), http://www.pbs.org/idealab/2013/02/beyond-foursquare-geolocation-services-proliferate-mature058/.

123. See Timothy B. Lee, *Here's How Phone Metadata Can Reveal Your Affairs, Abortions, and Other Secrets*, WASH. POST (Aug. 27, 2013), http://www.washingtonpost.com/blogs/the-switch/wp/2013/08/27/heres-how-phone-metadata-can-reveal-your-affairs-abortions-and-other-secrets; *see also* PCLOB REPORT, *supra* note 33, at 12 (stating "telephone calling records can reveal intimate details about a person's life, particularly when aggregated with other information and subjected to sophisticated computer analysis").

124. *See, e.g.*, Cyrus Farivar, *Volunteers in Metadata Study Called Gun Stores, Strip Clubs, and More*, ARS TECHNICA (Mar. 12, 2014), http://arstechnica.com/tech-policy/2014/03/volunteers-in-metadata-study-called-gun-stores-strip-clubs-and-more/.

125. WHITFIELD DIFFIE & SUSAN LANDAU, PRIVACY ON THE LINE: THE POLITICS OF WIRETAPPING AND ENCRYPTION 92 (1998). Traffic analysis is the process of intercepting communications and examining patterns in traffic data in order to gain intelligence. It can be performed even when the intercepted messages remain encrypted. *See* George Danezis & Richard Clayton, *Introducing Traffic Analysis*, DIGITAL PRIVACY: THEORY, TECHNOLOGIES, AND PRACTICES 95, 96 (Alessandro Acquisti et al. eds., 2008).

126. Tom Kellen, *Hiding in Plain View: Could Steganography Be a Terrorist Tool?*, SANS INSTITUTE, https://www.sans.org/reading-room/whitepapers/stenganography/hiding-plain-view-steganography-terrorist-tool-551 (last visited Mar. 17, 2014).

127. Peter Swire & Kenesa Ahmad, *'Going Dark' Versus a 'Golden Age for Surveillance*', CTR. FOR DEMOCRACY & TECH. (November 28, 2011), https://cdt.org/blog/%E2%80%98going-dark%E2%80%99-versus-a-%E2%80%98golden-age-for-surveillance%E2%80%99/.

Moreover, cybersecurity threats in particular can be embedded into all layers of a communication, regardless of the distinction between content and metadata. This means that protecting computers, networks and infrastructure against cyber risks requires monitoring not only of metadata but also of contents. Hence, with respect to the United States Computer Emergency Readiness Team's (US-CERT) Einstein Program, an intrusion detection system that monitors the network gateways of government agencies for cybersecurity risks, Dempsey writes that: "[t]he distinction between content and non-content is largely irrelevant to the Einstein debate, because Einstein undoubtedly captures and examines content, using a technique called deep-packet inspection."[128] This is corroborated by the Department of Homeland Security's privacy impact assessment for Einstein 3, which states that:

> DHS Office of Cybersecurity and Communications [CS&C] relies on signatures based on specific indicators that are known or suspected to be associated with malicious activity. While indicators will often be based on network traffic metadata, such as IP addresses, they may potentially be designed to match against any packet data, including the payload (the network traffic data). As such, E³A prevention capabilities may include deep packet inspection by ISPs.[129]

Paul Rosenzweig supports this approach, stating that "it would be an extremely poor rule that permitted screening of only non-content information for malware, as that would simply draw a map for malfeasant actors about how to avoid the intrusion detection systems."[130]

The Review Group recognized the weakness of the existing model, stating that "In a world of ever more complex technology, it is increasingly unclear whether the distinction between 'meta-data' and other information carries much weight."[131] It recommended that "the government should commission a study of the legal and policy options

128. James X. Dempsey, *Einstein 3.0: Liberty and Security Weigh in Favor of Private Sector Leadership*, *in* PATRIOTS DEBATE: CONTEMPORARY ISSUES IN NATIONAL SECURITY LAW (Harvey Rishikof, Stewart Baker & Bernard Horowitz, eds., 2012), *available at* http://www.americanbar.org/groups/public_services/law_national_security/patriot_debates2/the_book_online/ch6/ch6_ess2.html.

129. DEP'T HOMELAND SEC., PRIVACY IMPACT ASSESSMENT FOR EINSTEIN 3 - ACCELERATED (E³A), APR. 19, 2013, DHS/PIA/NPPD-027, *available at* http://www.dhs.gov/sites/default/files/publications/privacy/PIAs/PIA%20NPPD%20E3A%20130419%20FINAL%20signed.pdf [hereinafter *Einstein PIA*].

130. Paul Rosenzweig, *Providing for the Common Defense: The Government as Internet Protector*, *in* PATRIOTS DEBATE: CONTEMPORARY ISSUES IN NATIONAL SECURITY LAW (Harvey Rishikof, Stewart Baker & Bernard Horowitz, eds., 2012), *available at* http://www.americanbar.org/groups/public_services/law_national_security/patriot_debates2/the_book_online/ch6/ch6_ess1.html.

131. REVIEW GROUP, *supra* note 33, at 120–21.

for assessing the distinction between metadata and other types of information."[132]

The foregoing discussion supports a shift away from the traditional content-metadata dichotomy towards *a framework that assesses privacy risk based on the purpose of monitoring*. Different rules and procedures should apply to monitoring activities depending if they involve collection of evidence, intelligence gathering or cybersecurity defense. Where monitoring is restricted to cybersecurity defense, content can be conceptualized as a *container* for metadata, since its analysis is not intended to discern the "substance, purport, or meaning" of a communication. Rather it is meant to identify anomalies and signatures, including malware, viruses, Trojans, rootkits and phishing attacks (which are themselves non-content) that may be embedded in the content layer.[133] Hence, a machine would be reviewing the content of communication but only in search of suspicious metadata.[134] This monitoring could be analogized to a search performed by analysts who are non-English speakers, who can identify signatures of cybersecurity risks but are unable to comprehend the contents of the English-based communications that they sift through. Such analysts would technically be privy to the content of the communication but impervious to its "substance, purport and meaning." Clearly, they would be impotent if the purpose of the monitoring were the production of evidence or gathering of intelligence. Such a purpose would require application of different rules.

Advocating for a rule based on the purpose of monitoring should not be confused with support for a rule shifting privacy protections from the data collection to the data use stage. Over the past few years, several commentators have argued that privacy law should recalibrate to impose use, as opposed to collection-limitations.[135] This essay does not advocate wholesale data collection. On the contrary, it cautions against data retention and calls for analysis of data on the fly or upon very short

132. *Id.* Recommendation 6, at 25.

133. Milton Mueller & Andreas Kuehn, *Einstein on the Breach: Surveillance Technology, Cybersecurity and Organizational Change*, Paper Prepared for the 12thv Workshop on the Economics of Information Security (WEIS 2013), *available at* http://weis2013.econinfosec.org/papers/MuellerKuehnWEIS2013.pdf.

134. *Id.* at 12.

135. Fred H. Cate, Peter Cullen & Viktor Mayer-Schönberger, *Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines*, Mar. 2014, file:///Users/omer/Downloads/Data_Protection_Principles_for_the_21st_Century.pdf. *But cf.* Ann Cavoukian, Alexander Dix & Khaled El Emam, *The Unintended Consequences of Privacy Paternalism*, Mar. 2014, http://www.privacybydesign.ca/content/uploads/2014/03/pbd-privacy_paternalism.pdf.

periods (*e.g.*, milliseconds) of storage.[136]

Purpose-based rules for monitoring communications content and metadata can be based on two existing Supreme Court doctrines: the special needs doctrine and the contraband-specific doctrine.

### 1. The Special Needs Doctrine

The special needs doctrine establishes an exception to the Fourth Amendment warrant requirement, authorizing a reasonable government search without individualized suspicion where the immediate primary purpose of the search is to serve a special government interest other than gathering evidence for criminal law enforcement purposes.[137] It has been used to authorize searches in various contexts, including at border crossings[138] and highway checkpoints,[139] as well as random checks of government employees.[140]

In a long line of cases, the Supreme Court authorized the government to conduct suspicionless searches at the international border or its functional equivalent.[141] One commentator argued that "the sovereign's right to protect itself at the digital international border should be at least as coextensive as its right to protect itself in the physical world."[142] Similarly, the Supreme Court sanctioned the use of police checkpoints to scan drivers for signs of intoxication[143] or detect the presence of illegal immigrants.[144] Immigration checkpoint stops were held to be "reasonable" under the Fourth Amendment because the state's interest in detecting the presence of illegal aliens outweighed the limited intrusion on an individual's privacy that was caused by the checkpoints.

136. *Infra* notes 178–179 and accompanying text; *see also* PCLOB REPORT, *supra* note 33, at 10 (stating "Section 215 is designed to enable the FBI to acquire records that a business has in its possession, as part of an FBI investigation, when those records are relevant to the investigation. Yet the operation of the NSA's bulk telephone records program bears almost no resemblance to that description") and 13 ("detailed rules currently in place limit the NSA's use of the telephone records it collects…. But in our view, they cannot fully ameliorate the implications for privacy, speech, and association that follow from the government's ongoing collection of virtually all telephone records of every American").

137. O'Connor v. Ortega, 480 U.S. 709 (1987); Skinner v. Railway Labor Executives Association, 489 U.S. 602 (1989).

138. United States v. Place, 462 U.S. 696 (1983).

139. Michigan Dep't of State Police v. Sitz, 496 U.S. 444 (1990).

140. Skinner v. Railway Labor Executives' Ass'n, 489 U.S. 602 (1989); Treasury Employees v. von Raab, 489 U.S. 656 (1989).

141. *See, e.g.*, United States v. Ramsey, 431 U.S. 606, 616 (1977); United States v. Montoyo de Hernandez, 473 U.S. 531, 538 (1985).

142. Scott J. Glick, *Virtual Checkpoints and Cyber-Terry Stops: Digital Scans To Protect the Nation's Critical Infrastructure and Key Resources*, 6 J. NAT'L SEC. L. & POL'Y 1, (2012).

143. Michigan Dep't of State Police v. Sitz, 496 U.S. 444 (1990).

144. United States v. Montoyo de Hernandez, 473 U.S. 531 (1985).

In all of these cases, "a search remains a valid administrative search only so long as the scope of the search is permissibly narrow; once a search is conducted for a criminal investigatory purpose, it can no longer be justified under an administrative search rationale."[145]

The Snowden revelations have shown that in a series of decisions, the FISC extended the "special needs" doctrine into the realm of national security electronic monitoring.[146] In one case, the FISC held that:

> "The question. . . is whether the reasoning of the special needs cases applies by analogy to justify a foreign intelligence exception to the warrant requirement . . . Applying principles derived from the special needs cases, we conclude that this type of foreign intelligence surveillance possesses characteristics that qualify it for such an exception.[147]

While controversial, this decision demonstrates the proclivity of judges to recognize the rationale of an administrative search in a cyberspace.

The "special needs" exception has also been used to deploy body scanners at airports.[148] When passing through a body scanner, the human body is presented as an instrumental container of contraband or explosives.[149] If passengers were required to parade naked under the gaze of a TSA officer, they would no doubt experience an invasion of privacy. The scan is therefore structured both technologically and operationally to avert any direct interaction between an officer and an undressed human body. Tirosh and Birnhack explain:

> Technologically, the scanners blur faces, and now use only a generic outline of a human body. Operationally, the agent reviewing the image sits in a remote location and does not see the passenger. . . Discursively, the language that is applied to describe scanning presents it as an automatic, anonymized, universal, neutral, routine,

145. United States v. $124,570 U.S. Currency, 873 F.2d 1240, 1246 (9th Cir. 1989).

146. Eric Lichtblau, *In Secret, Court Vastly Broadens Powers of N.S.A.*, NY TIMES (July 6, 2013), http://www.nytimes.com/2013/07/07/us/in-secret-court-vastly-broadens-powers-of-nsa.html.

147. *In re* Directives Pursuant to Section 105B of the Foreign Intelligence Surveillance Act, 551 F.3d 1004, 1011–1012 (FISA Ct. Rev. 2008); Cassidy v. Chertoff, 471 F.3d 67, 82 (2d Cir. 2006).

148. Alexander A. Reinert, *Revisiting "Special Needs" Theory Via Airport Searches*, 106 NW. L. REV. COLL. 207 (2012), *available at* http://www.law.northwestern.edu/lawreview/colloquy/2012/2/LRColl2012n2Reinert.pdf.

149. Yofi Tirosh & Michael Birnhack, *Naked in Front of the Machine: Does Airport Scanning Violate Privacy?*, 74 OHIO ST. L. J. 1263 (2013).

and professional process.[150]

In short, the officer is blind to the human body under scrutiny and instead focuses only on any illicit material carried by the passenger.[151] Similarly, in the cybersecurity context, the monitoring machine would ignore the content of a communication, focusing only on any signature of cyber attack.

For this approach to work, Chinese walls must be erected between communication monitoring for cybersecurity and law enforcement or foreign intelligence purposes. "The Supreme Court has repeatedly emphasized the importance of keeping criminal investigatory motives from coloring administrative searches."[152] Although the FISC stated that "[a] surveillance with a foreign intelligence purpose often will have some ancillary criminal-law purpose,"[153] the diffusion of data gleaned through cybersecurity monitoring into the criminal system would undercut constitutional protections and criminal procedure law. Such a separation of objectives will no doubt raise challenging dilemmas: should information parsed through cybersecurity monitoring be used to apprehend and indict terrorists or child pornographers? Should it be used to defuse a "ticking bomb" threat? Such questions exceed the scope of this Essay. Suffice it to say, that any repurposing of data would undermine the legitimacy of cybersecurity operations.

2. The Contraband-specific Doctrine

Like the special needs doctrine, the contraband-specific doctrine has developed in Supreme Court rulings concerning dog sniff searches by the police. In *Illinois v. Caballes*, the Court held that a narcotics-sniffing dog search conducted during a routine traffic stop did not violate the Fourth Amendment, given that the search "reveals no information other than the location of a substance that no individual has any right to possess".[154] Similarly, in *Place*, the Court held that a police dog sniff is a *sui generis* search since "[there is] no other investigative technique that is so limited both in the manner in which the information is obtained and in the content of the information revealed by the procedure."[155] With a dog sniff, the police do not violate passengers' privacy by digging through their personal belongings; instead they conduct an external review to

150. *Id.* at 1269.
151. *Id.*
152. Camara v. Municipal Court, 387 U.S. 523 (1967).
153. *In re* Sealed Case, 310 F.3d 717, 743 (FISA Ct. Rev. 2002).
154. Illinois v. Caballes, 543 U.S. 405, 406 (2005).
155. United States v. Place, 462 U.S. 696 (1983).

detect just the "presence or absence of narcotics."[156]

Indeed, a contraband-specific focused search is more protective of privacy than potential alternatives. As mentioned above, the Supreme Court has described a canine sniff as "unobtrusive," as compared to an officer digging through luggage, on the theory that "the manner in which information is obtained through this investigative technique is much less intrusive than a typical search."[157]

A similar analysis can apply to a cybersecurity scan of communication contents. Such contents are of course private and may be highly sensitive, like the content of a passenger's luggage, which could include medication, underwear, intimate items, pornographic materials, legal documentation, etc. But just as the narcotics sniffing dog flags only suspicious items, so does the mechanical sniffer scrutinize only communications susceptible to cybersecurity threats. Indeed, the algorithmic scanning of massive amounts of communications can forestall more intrusive monitoring techniques, which entail a deep dive into the contents of a suspect's communications.

It is important to note that automated monitoring is laden with a persistent risk of false positives, that is, the flagging of an innocent communication as suspect and subjecting it to additional scrutiny.[158] Individuals who are party to such communications will no doubt experience an invasion of privacy, as human analysts will eventually closely review their case. Any system calibrated to prevent false negatives will inevitably produce false positives. To address the privacy problems that arise in such cases, robust mechanisms of privacy by design are required.

Despite this risk, such false positives should not be assessed in a vacuum. Given the nature of cybersecurity and national security risks, it is not that no privacy intrusions would take place absent a program for communications monitoring. Rather, the government would have to come up with alternative systems to protect critical infrastructure and national security and apply different, perhaps more intrusive, tools. As explained in the Supreme Court's *Place* and *Jacobson* decisions, the intrusiveness of a system should be assessed against other mechanisms for search. For example, if body scanners are removed from airports in the name of privacy, civil liberty advocates could conceivably score a Pyrrhic victory, with more intrusive security measures introduced instead.

156. *Id.* at 707.

157. *Id.*; *see also* United States v. Jacobsen, 466 U.S. 109 (1984).

158. *See, e.g.*, MHR Khouzani, Soumya Sen & Ness B. Shroff, *An Economic Analysis of Regulating Security Investments in the Internet*, *in* PROC. OF IEEE INFOCOM, TURIN, ITALY, 2013, *available at* http://www.tc.umn.edu/~ssen/papers/Security-Infocom2013.pdf.

### C. *Privacy by Design*

Privacy by design has become a catchphrase in privacy policymaking that means different things to different people. Initially coined by Ontario Privacy Commissioner Ann Cavoukian,[159] it has been endorsed by policymakers in the United States,[160] the European Union[161] and OECD.[162] This essay proposes several parameters for analysis of the privacy impact of a communications monitoring program. By minimizing the privacy impact of a monitoring program along each of these parameters, policymakers can embed privacy in the program's design. The parameters are categorized into three groups: *organizational parameters*, demarcating the scope and nature of a monitoring program; *technological parameters*, such as data anonymization, retention limits and persistent audit logs; and *legal parameters*, such as oversight by the judicial and legislative branches, appointment of a chief privacy officer and sanctions for violations.

#### 1. Organizational Parameters

##### *Who Monitors?*

A cybersecurity system that is managed by the *private sector* will raise fewer privacy concerns than one overseen by the *government*. The monopoly that the state holds over legitimate use of force[163] creates a qualitative difference between power allocated to the government and to the private sector. If worse come to worst, all that Google can do to a user is target him or her with bothersome ads, whereas the government can take away a citizen's liberty or, in some cases, life.[164] Similarly, a system deployed by the government will generate fewer privacy

159. *See About PbD*, PbD, http://www.privacybydesign.ca/index.php/about-pbd/ (last visited Mar. 17, 2014).

160. FED. TRADE COMM'N, *supra* note 97, at 22–34.

161. DRAFT REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL ON THE PROTECTION OF INDIVIDUALS WITH REGARD TO THE PROCESSING OF PERSONAL DATA AND ON THE FREE MOVEMENT OF SUCH DATA (GENERAL DATA PROTECTION REGULATION) (Oct. 2013), *available at* http://www.janalbrecht.eu/fileadmin/material/Dokumente/DPR-Regulation-inofficial-consolidated-LIBE.pdf.

162. ORG. FOR ECON. CO-OPERATION & DEV., RECOMMENDATION OF THE COUNCIL CONCERNING GUIDELINES GOVERNING THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA (2013) [C(80)58/FINAL, AS AMENDED ON 11 JULY 2013 BY C(2013)79], *available at* http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf.

163. *See* Max Weber, *Politics as a Vocation*, *in* THE VOCATION LECTURES (David Owen, ed., Hackett Publishing, 2004).

164. PCLOB REPORT, *supra* note 33, at 12 ("With its powers of compulsion and criminal prosecution, the government poses unique threats to privacy when it collects data on its own citizens").

concerns if managed by a *civilian agency* rather than a military, intelligence, or *national security agency*. This results from the lack of transparency inherent in the operation of national security agencies as well as the fear of creating omnipotent data-rich government departments such as the East German Stasi.

A disposition to take mass surveillance out of the hands of government is evident in the Review Group's recommendation that "legislation should be enacted that terminates the storage of bulk telephony meta-data by the government under section 215, and transitions as soon as reasonably possible to a system in which such meta-data is held instead either by private providers or by a private third party."[165] This recommendation, as opposed to some of the others made by the Review Group, was adopted by President Obama in his speech about the NSA revelations: "I am therefore ordering a transition that will end the Section 215 bulk metadata program as it currently exists, and establish a mechanism that preserves the capabilities we need without the government holding this bulk meta-data."[166] Some commentators argue that the private sector is not only less risk-prone than the government but also better placed to address cybersecurity risks given its agility and diverse knowledge base.[167] Others fear that private sector entities are ill equipped to securely manage such rich and sensitive databases, and would be tempted to use the information they store for various business purposes.[168] They posit that in some cases, it may be safer to keep the data in the government's hands.[169]

165. REVIEW GROUP, *supra* note 33, Recommendation 5, at p. 25.

166. Remarks of President Barack Obama, Results of our Signals Intelligence Review, Jan. 17, 2014, Washington, D.C. [hereinafter *NSA Speech*], *available at* http://www.lawfareblog.com/2014/01/text-of-the-presidents-remarks-on-nsa-and-surveillance/#.UtyugxAo671.

167. See Dempsey, *supra* note 128 ("In my view, the private sector is both more agile and more knowledgeable in key respects about its systems than the federal government could ever be. To the extent that the federal government has some specialized knowledge that would be helpful to the private sector, the goal of policy should be to transfer that knowledge to the private sector in a way that is both secure and useful. Leaving the main responsibility for protecting private sector networks in the hands of the private sector will not only be most effective from a security standpoint, but it will also have significant civil liberties benefits as well.").

168. Bruce Schneier, *Let the NSA Keep Hold of the Data*, SLATE, Feb. 14, 2014 3:03 PM, http://www.slate.com/articles/technology/future_tense/2014/02/nsa_surveillance_metadata_the_government_not_private_companies_should_store.html; *also see* Jack Goldsmith, *Cole and Lederman, and Morell, on Review Group Report*, LAWFARE Dec. 23, 2013, http://www.lawfareblog.com/2013/12/cole-and-lederman-and-morell-on-review-group-report.

169. *Id.* (concluding: "If the corporations are storing the data already—for some business purpose—then the answer is easy: Only they should store it. If the corporations are not already storing the data, then—on balance—it's safer for the NSA to store the data. And in many cases, the right answer is for no one to store the data. It should be deleted because keeping it

*Monitor Whom?*

Not only the identity of the entity doing the monitoring but also the scope of the monitored systems is an important parameter from a privacy standpoint. A monitoring program that only protects *government assets* is far less sensitive than one that extends to the private sector. Other options include monitoring programs that apply to government assets as well as to privately owned *critical infrastructure*, such as utilities and telecommunication systems, or private sector businesses that *voluntarily* choose to collaborate with the government.[170] In its privacy impact assessment of the Einstein 3 system, the Department of Homeland Security (DHS) explains the expanding scope of the Einstein programs:

> In brief, EINSTEIN 1 analyzes network flow records and EINSTEIN 2 detects and alerts to known or suspected cyber threats using Intrusion Detection Systems (IDS) technology. EINSTEIN 2's network IDS technology uses custom signatures, based upon known or suspected cyber threats within federal network traffic. [Einstein 3] supplements EINSTEIN 2 by adding additional intrusion prevention capabilities and enabling ISPs, under the direction of DHS, to detect and block known or suspected cyber threats using indicators.[171]

Notably, the DHS states that the legal basis for deploying Einstein in its various iterations is consent, not only of government personnel but also of relevant third parties, since "[o]nce an individual decides to communicate with a participating agency electronically, the network traffic is subject to computer security efforts of CS&C, including in this case E³A, in addition to any individual computer security programs the agency might have in place."[172]

2. Technological Parameters

*De-identification and Limited Retention*

Anonymization, or de-identification, the removal or masking of personal data from a dataset, remains one of the most useful privacy enhancing mechanisms. Over the past few years, technology[173] and

makes us all less secure.")

170. Consider, for example, the Defense Industrial Base Cybersecurity/Information Assurance Program (DIB CS/IA), *DIB CS/IA Program*, DEP'T OF DEFENSE, http://dibnet.dod.mil (last visited Mar. 17, 2014).

171. *Einstein PIA*, *supra* note 129.

172. *Id.* at 19.

173. Arvind Narayanan, *33 Bits of Entropy: The End of Anonymous Data and What to do About it*, 33 BITS OF ENTROPY, http://33bits.org (last visited June 1, 2014); *but cf.* Khaled El Emam, GUIDE TO THE DE-IDENTIFICATION OF PERSONAL HEALTH INFORMATION (2013).

policy experts[174] have engaged in bitter arguments about the robustness of de-identification in light of escalating re-identification attacks.[175] Regardless of the outcome of this ongoing debate, few would argue that de-identification does not at least *mitigate* privacy risk. De-identification has been advocated in a host of privacy policy documents in the commercial sector, including by the White House[176] and the FTC.[177]

In addition to de-identification, the administrators of monitoring programs should implement data minimization and limit data retention periods. Indeed, monitoring data on the fly *without any data retention* or retention for limited time periods such as milliseconds required to conduct computational analysis would greatly reduce privacy risks. In its privacy impact assessment of Einstein 3, the DHS alludes to de-identification and data minimization, noting that "CS&C Standard Operating Procedures [SOPs] and information handling guidelines require CS&C cybersecurity analysts to minimize (*i.e.*, overwrite, redact, or replace) personally identifiable information [PII] data that is not necessary to understand the cyber threat."[178]

One of the recommendations of the Review Group, although aspirational, advocates data analysis on the fly without any retention:

> It might reduce budgetary costs and political risk if technical collection agencies could make use of artificial intelligence software that could be launched onto networks and would be able to determine in real time what precise information packets should be collected. Such smart software would be making the sorting decision online, as distinguished from the current situation in which vast amounts of data are swept up and the sorting is done after it has been copied on to

174. Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010); *but cf.* Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J. L. & TECH. 1 (2011).

175. Yianni Lagos & Jules Polonetsky, *Public vs. Nonpublic Data: The Benefits of Administrative Control*, 66 STAN. L. REV. ONLINE 103 (2013).

176. THE WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY 21 (Feb. 2012), http://www.whitehouse.gov/sites/default/files/privacy-final.pdf ("Companies should collect only as much personal data as they need to accomplish purposes specified under the Respect for Context principle. Companies should securely dispose of or de-identify personal data once they no longer need it, unless they are under a legal obligation to do otherwise.").

177. FED. TRADE COMM'N, *supra* note 97, at iv ("data is not "reasonably linkable" to the extent that a company: (1) takes reasonable measures to ensure that the data is de-identified; (2) publicly commits not to try to re-identify the data; and (3) contractually prohibits downstream recipients from trying to re-identify the data.")

178. *Einstein PIA*, *supra* note 129, at 9.

data storages systems.[179]

At the same time, de-identification and limited retention are inhibited by the realization that engaging cybersecurity risks may require analysts to re-identify personal data in order to reach suspects of attacks. In addition, technical experts may argue that analysis of immense amounts of communications data in real time is not technologically feasible absent some period of data retention.

*Audit Logs*

An additional technology, which mitigates potential abuses of monitoring programs, is immutable audit logs. A Markle Foundation task force examining information sharing in the digital age noted, "[t]he ability to maintain tamper-resistant logs of user activity on the network can increase security, build trust among users, ensure compliance with relevant policies and guidelines, and improve transparency and the ability to perform oversight by appropriate stakeholders outside the system."[180] Audit logs can record everything from login attempts to specific user search queries to user views of individual records. They help operationalize concepts of accountability and ensure compliance with the law and applicable policies.

3. Legal Parameters

*Oversight Mechanisms*

As has become abundantly clear from the maelstrom surrounding NSA surveillance, sound legal safeguards and processes are vital to temper abuse of government power. Robust oversight processes, not only internal within an intelligence organization or integral to the executive branch (*e.g.*, the Attorney General) but also by the legislature and judiciary, must be put in place to curb mass surveillance. The Review Group devoted many of its recommendations to putting in place such processes, including oversight by senior intelligence officials[181] and

179. *See, e.g.*, REVIEW GROUP, *supra* note 33, Recommendation 20, at 173.

180. *Implementing a Trusted Information Sharing Environment, Using Immutable Audit Logs to Increase Security, Trust and Accountability*, MERKLE FOUNDATION (February 2006), http://www.markle.org/sites/default/files/nstf_IAL_020906.pdf.

181. *See, e.g.*, REVIEW GROUP, *supra* note 33, Recommendation 18 (the Director of National Intelligence to establish a mechanism to monitor the collection and dissemination activities of the Intelligence Community to ensure they are consistent with the determinations of senior policymakers); Recommendation 24 (separating the role of the head of the military unit, US Cyber Command, and the Director of the National Security Agency); Recommendation 25 (spinning off the NSA's Information Assurance Directorate to become a

newly created organizational structures,[182] as well as by the President,[183] Congress,[184] the courts,[185] and the public at large.[186] Of particular importance is making information available to the general public, to avert the eerie specter of a Kafkaesque bureaucracy operating in the shadows and pursuing an opaque agenda.[187]

In his NSA Speech, the President accepted the Review Group's recommendation to require a judicial order prior to any individual query on the government's communication database, stating "I have directed the Attorney General to work with the Foreign Intelligence Surveillance Court so that during this transition period, the database can be queried only after a judicial finding, or in a true emergency.[188]

One complication with fully automated monitoring processes is finding the right juncture at which judicial oversight can interject. Yet automation should not forestall legal safeguards. Whenever an automated process zooms in to focus on an individual suspect, due process concerns arise. In addition, an automated process will inevitably produce false

separate agency within the Department of Defense).

182. *See, e.g.*, *id.*, Recommendation 26 (the creation of a privacy and civil liberties policy official located both in the National Security Staff and the Office of Management and Budget); Recommendation 27 (strengthening the Privacy and Civil Liberties Oversight Board and authorizing it to receive whistle-blower complaints from employees of the intelligence agencies); Recommendation 28 (create the position of Public Interest Advocate to represent privacy and civil liberties interests before the FISC).

183. *See, e.g., id.*, Recommendation 16 (the President to create a new process requiring high-level approval of all sensitive intelligence requirements and the methods the Intelligence Community will use to meet them).

184. *See, e.g.*, *id.*, Recommendation 7, 18 (regular reports to Congress on use of electronic surveillance programs); Recommendation 22 (the Director of the National Security Agency to be a Senate-confirmed position, possibly held by a civilian.)

185. *See, e.g.*, *id.*, Recommendation 1 ("We believe that, as a matter of sound public policy, it is advisable for a neutral and detached judge, rather than a government investigator engaged in the 'competitive enterprise' of ferreting out suspected terrorists, to make the critical determination whether the government has reasonable grounds for intruding upon the legitimate privacy interests of any particular individual or organization" – at p. 88); Recommendation 2 (judicial authorization for issuance of National Security Letters); Recommendation 8 (judicial authorization required for non-disclosure orders); *see also* PCLOB REPORT, *supra* note 33, Recommendation 2 at 17.

186. *See, e.g.*, *id.*, Recommendations 9, 10 (public disclose by companies of general information about the number of orders they received, the categories of information they produced, and the number of users impacted); Recommendation 28 (enhanced transparency of FISC decisions); *see also* PCLOB REPORT, *supra* note 33, at 15 (stating "The Board believes that the government must take the initiative and formulate long-term solutions that promote greater transparency for government surveillance policies more generally, in order to inform public debate on technology, national security, and civil liberties going beyond the current controversy.").

187. *See, e.g.*, Michael Winship, *Snowden's Legal Counsel: Forget About Orwell, Worry About Kafka*, MOYERS & CO. (Mar. 11, 2014), http://billmoyers.com/2014/03/11/our-chat-with-edward-snowdens-legal-counsel.

188. *NSA Speech*, *supra* note 166.

positives requiring individual protections under the law. This means that whenever an individual suspect is singled out for further investigation, legal process must be triggered.

### *Accountability*

In addition to mechanisms of legal oversight, monitoring programs require the creation of operational accountability processes within intelligence and national security agencies. Accountability was one of the original fair information practice principles, first iterated by the 1980 OECD Privacy Guidelines.[189] Accountability has gained traction in recent years as a practically oriented requirement to operationalize privacy policies.[190] Accountability measures include the appointment of a Chief Privacy Officer, who is charged with overseeing a privacy program and coordinating with employees embedded in different parts of an organization.[191] Indeed, as a result of the Snowden fallout, the NSA has recently appointed its first ever Chief Privacy Officer, herself a former Deputy Chief Privacy Officer at the DHS, one of the government's pioneering privacy programs.[192] Additional measures include conducting privacy impact assessments prior to the implementation of new systems or programs, and imposing individual liability, including potential fines and penalties, for violation of organizational privacy rules.

## CONCLUSION

This essay proposes new parameters for surveillance regulation based on an analysis of privacy harm that is grounded in the current state of play of information technologies. It explores new concepts such as *strictly automated processing* and *content as a container for metadata*. It suggests that the measurement of privacy harm must transcend traditional legal categories and instead focus on factors such as the identity of the entity performing the monitoring or retention of information (*e.g.*, military intelligence, civilian agency, private sector); the protected zone (*e.g.*, military assets, critical infrastructure, private

189. ORG. FOR ECON. CO-OPERATION & DEV., GUIDELINES GOVERNING THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA, OECD DOC. C(80)58/FINAL (Sept. 23, 1980), http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

190. GENERAL DATA PROTECTION REGULATION, *supra* note 161.

191. Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247, 253 (2011).

192. Al Kamen, *The NSA Has A New, First Time Ever, Privacy Officer*, WASH. POST (Jan. 28, 2014), http://www.washingtonpost.com/blogs/in-the-loop/wp/2014/01/28/the-nsa-has-a-new-first-time-ever-privacy-officer.

sector); the purposes of monitoring (*e.g.*, foreign intelligence, law enforcement, cybersecurity); the degree of automation (*e.g.*, the calibration of acceptable levels of false positives or negatives); the retention periods (*e.g.*, real time interception, very short term retention, medium or long term retention); and deployment of mechanisms of privacy by design including organizational, technological and legal measures.