A TWENTY-FIRST CENTURY FRAMEWORK FOR DIGITAL PRIVACY

WHITE PAPER SERIES

Policing and The Cloud

CHRISTOPHER SLOBOGIN

NATIONAL CONSTITUTION CENTER

Policing and The Cloud

Christopher Slobogin*

It is now a commonplace that virtually everything we do is memorialized on databases—databases which, for brevity's sake, this paper will refer to as The Cloud, despite the fact that not all the data this paper discusses is found there. These databases—the servers of Google, Netflix, and Apple; the memory banks of phones, closed circuit cameras, "smart cars," and satellites; the computers in commercial establishments and government agencies—track an astonishing range of our intimate daily activities, including financial transactions, Internet connections, travel routes, tax information, and medical treatment, as well as more prosaic matters such as employment and residence history, utility usage, and car malfunctions. The question addressed here is when the government should be able to gain access to this wealth of personal information for law enforcement and national security purposes.

In the United States, answering that question requires consulting a welter of statutes and a few Supreme Court decisions. For instance, when the government wants to access information stored on a computer or found in texts or emails, federal and state laws usually require a warrant, issued by a judge who has found probable cause that the communication will lead to evidence of wrongdoing.[1] However, if officials want an already opened communication or one that has been on a server for over 180 days, then they may only need to show that the communication is "relevant" to an investigation, a much lower standard than probable cause, albeit an assertion that is challengeable by the target, as occurs with an ordinary subpoena.[2] And if the communication sits on a "private" server (for instance, a private university or employer), no court process is required.[3]

When law enforcement officials seek records outside the communications context, a wide array of statutes may be applicable. As a general matter, bank, educational, and even medical records can be obtained with a mere subpoena, which the target often does not find out about unless and until prosecution occurs.[4] In a host of other situations, such as accessing camera footage or obtaining data about credit card purchases or past travel routes, most jurisdictions do not require police to follow any judicial process, but rather allow them to obtain the information at their discretion and that of the data holders.[5]

---

* The author would like to thank Stephen Henderson, Wayne Logan, Scott Sundby and Robert Weisberg for their comments on earlier versions of this chapter.

[1] *See, e.g.*, 18 U.S.C. §§ 2511 & 2518.

[2] 18 U.S.C. § 2703(a), (b) (1) (B). If the information is stored on a private server (e.g., one run by a business) no court process is required. 18 U.S. C. § 2711(2) (defining remote computing service). As of April 30, 2017, the House of Representatives had unanimously voted to repeal this provision and require a warrant but the Senate had yet to vote.

[3] Quon v. Arch Wireless Operating Co. Inc., 445 F.Supp.2d 1116, 1130 (2006).

[4] For a summary, *see* CHRISTOPHER SLOBOGIN, PRIVACY AT RISK: THE NEW GOVERNMENT SURVEILLANCE AND THE FOURTH AMENDMENT 174-175 (2007).

[5] For instance, under the federal Privacy Act, if the Department of Justice wants data from another federal agency, it need merely make a written request. 5 U.S.C. § 552a(b)(7). Even that much is not required when the data is obtained through a private databroker, since the data is not coming directly from an agency. Chris Jay Hoofnagle,

In theory, the Constitution, and in particular, the Fourth Amendment, could have something to say about all of this. The Fourth Amendment requires that the government act reasonably when it engages in a "search" or "seizure," and the courts have held that, for many types of searches, this reasonableness requirement can only be met with a warrant. However, this requirement only applies to government actions that are considered "searches." The Supreme Court has defined that word very narrowly, to encompass only those actions that infringe "reasonable expectations of privacy" or that involve some type of physical intrusion.[6] Most relevant here are the Court's decisions holding that expecting constitutional protection from government acquisition of information surrendered to third parties—whether they be internet service providers, banks, or phone companies—is not reasonable, since we "assume the risk" that those third parties will decide to give that information to the government.[7] As discussed below, this "third party" doctrine has seen some erosion in recent years, but it remains the reason that, other than when access to the content of communications is involved,[8] the Fourth Amendment has very little impact on the government's ability to obtain information, even when it relies on technology to do so.

While many have inveighed against the laxness of both statutory and constitutional law, the most popular counter-proposal—that Cloud access by the government should require a judicial warrant—has problems of its own. Conceptually, a warrant requirement glosses over the intuition that a large number of situations, while involving a viable privacy claim vis-à-vis the government, do not merit the full protection of a warrant. Practically, it would handcuff legitimate government efforts to nab terrorists and criminals. A more nuanced approach is necessary.

That approach should begin with an assessment of the varying motivations that drive the government's use of The Cloud. Cloud-searches can come in at least five different guises: suspect-driven, profile-driven, event-driven, program-driven, or volunteer-driven. Some Cloud access by the state is aimed at getting as much information as possible about individuals suspected of wrongdoing. Other efforts do not start with a particular suspect, but rather with a profile of a hypothetical suspect, purportedly depicting the characteristics of those who have committed or will commit a particular sort of crime. A third type of Cloud-search starts neither with a suspect nor a suspect profile but with an event—usually a crime—and tries to figure out, through location and related information, who might be involved. Fourth, so as to have the information needed for suspect-, profile-, and event-driven operations at the ready, government

---

Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement, 29 N.C. J. INT'L & COMM. REG. 595, 623 (2003).

[6] Kyllo v. United States, 533 U.S. 27, 33 (2001); Florida v. Jardines, 133 S.Ct. 1409, 1414 (2013) (indicating that the expectation of privacy test is supplemented by inquiry into whether the government "engage[s] in [a] physical intrusion of a constitutionally protected area").

[7] *See, e.g.*, United States v. Miller, 425 U.S. 435, 443 (1976) ("This Court has held repeatedly that the Fourth Amendment does not prohibit the obtaining of information revealed to a third party and conveyed by him to Government authorities, even if the information is revealed on the assumption that it will be used only for a limited purpose and the confidence placed in the third party will not be betrayed."); Smith v. Maryland, 442 U.S. 735, 743-44 (1979) (same holding with respect to phone numbers dialed).

[8] *See, e.g.*, United States v. Warshak, 631 F.3d 266 (6th Cir. 2010) (finding that the Fourth Amendment requires a warrant to obtain stored emails).

might initiate data collection programs. Finally, the government also relies on citizens to come forward on their accord when they find incriminating information about another person.

Each of these Cloud-based endeavors are distinct from the other four. Each calls for a different regulatory regime. Below is a sketch of what those regimes might look like. While they borrow from Fourth Amendment jurisprudence, the principles developed here fill a void because, to date, that jurisprudence has had little to say about Cloud searches. Until the Supreme Court weighs in, policymakers are working pretty much on a clean slate in this area.

I. Suspect-Driven Cloud Access—Proportionality

Assume the police receive an anonymous phone call from a female claiming that John Slade, a fifth grade public school teacher, is also a drug dealer. In investigating this claim, police might want to obtain Slade's phone records to see if he's called known drug dealers, gang members, or drug users. They might also seek access to his bank records to discover whether the amount of money he deposits is consistent with his job as a school teacher. Additionally, police might like to find out from GPS records and drone and camera feeds if Slade frequents areas of town where drugs are routinely sold.

Under current Fourth Amendment and statutory law, none of these policing moves requires a warrant or probable cause and, depending on the jurisdiction, some of it may not even require a subpoena. That lack of regulation is abetted by the Supreme Court's assertion that expecting privacy in information surrendered to a third party or in activities carried out in public is unreasonable.[9] Yet most people surveyed on these matters come to a quite different conclusion, ranking perusal of their bank and phone records, for instance, as comparable to search of a bedroom, and ranking location tracking as similar in invasiveness to a frisk.[10] On a more philosophical plane, scholars argue that allowing the government to invade The Cloud so easily offends not only privacy, but autonomy and dignity.[11] They also claim it chills citizens' rights to expression and association, and creates huge potential for abuse; after all, Knowledge—which The Cloud provides in troves—is Power.[12]

---

[9] *See supra* note 7 and United States v. Knotts, 460 U.S. 276, 281 (1983) (holding that there is no expectation of privacy on public thoroughfares).

[10] *See* tables depicting research in SLOBOGIN, *supra* note 4, at 112 (compare items 14 and 16); 184 (compare items 24 and 25). This research has been replicated in Christine S. Scott-Hayward, Henry F. Fradella & Ryan G. Fischer, *Does Privacy Require Secrecy?: Societal Expectations of Privacy in the Digital Age*, 43 AM. J. CRIM. L. 19 (2015) and Jeremy E. Blumenthal, Meera Adya & Jacqueline Mogle, *The Multiple Dimensions of Privacy: Testing "Lay" Expectations of Privacy*, 11 U. PA. J. CONST. L. 331, 345 (2009).

[11] *See, e.g.*, David Lametti, *The Cloud: Boundless Digital Potential or Enclosure 3.0*, 17 VA. J.L. & TECH. 190, 196 (2012) ("we may be witnessing another round of 'enclosure' in Cloud space that might have serious deleterious effects on what we have come to expect in the digital age: autonomy, exchange, spontaneity, and creativity, and all at a lightning pace.").

[12] *See, e.g.*, Daniel Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1461 (2001) ("The problem with databases is . . . a problem that involves power and the effects of our relationship with public and private bureaucracy—our inability to participate meaningfully in the collection and use of our personal information.").

The Supreme Court itself has begun to recognize these concerns. In *Riley v. California*,[13] despite centuries-old precedent permitting suspicionless searches of any item found on an arrested individual, it required a warrant for a search of a cell phone of an arrestee, in recognition of the fact that "the cell phone collects in one place many distinct types of information—an address, a note, a prescription, a bank statement, a video—that reveal much more in combination than any isolated record."[14] In *United States v. Jones*,[15] five members of the Court concluded that a Fourth Amendment search occurs when the police engage in "prolonged" tracking of a vehicle using GPS signals. While neither *Riley* nor *Jones* involved Cloud access, Justice Sotomayor may have summed up where the Court is going when she stated in her concurring opinion in *Jones* that "it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties. This approach is ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks."[16]

On this view, government would not be foreclosed from perusing, at its discretion, blogs, tweets, public records, and other sources that are clearly meant to be consumed by the public. But it would prohibit police from accessing, in the absence of justification, non-public Cloud data people generate when they engage in "mundane tasks" like communicating with their friends, banking, and shopping. It would also prohibit suspicionless access to tracking data about everyday travels that the average person undertakes on the assumption of practical anonymity.

In short, there is a strong case to be made for requiring the government to demonstrate it has good reason to go after Cloud-based information about a particular person that is not readily available in public fora. Then the question becomes how good that reason must be. Normally, the Fourth Amendment requires that a search be based on probable cause, which amounts to a "fair probability" that a search will discover evidence of crime.[17] Return to the investigation of Slade. If the caller had identified herself and provided detail about Slade's drug deals, police would have had probable cause and grounds for a full-scale digital search. But recall that, in fact, the caller was anonymous and simply said Slade was selling drugs, thus making it difficult to dismiss the possibility that she was a disgruntled student or a spurned lover. Under Supreme Court caselaw, that call, by itself, would not permit a traditional search.[18]

But suppose instead that the call, although anonymous, provides detail about Slade's next drug deal. While, by itself, this would not be enough for probable cause, its predictive quality does provide an additional indication of reliability.[19] In that intermediate situation, police might arguably have "reasonable suspicion" (a lesser level of cause but one that nonetheless requires an

[13] 134 S.Ct. 2473 (2014).
[14] *Id*. at 2489.
[15] 132 S.Ct. 945 (2012).
[16] *Id.* at 947.
[17] *See* WAYNE R. LAFAVE ET AL., 2 CRIMINAL PROCEDURE 114-115 (3d ed. 2007).
[18] Florida v. J.L., 529 U.S. 266 (2000) (holding unconstitutional a frisk based on an anonymous phone call stating that the defendant would be standing on a street corner wearing certain clothing with a gun on his person).
[19] *Cf.* Illinois v. Gates, 462 U.S. 213 (1983) (holding police had probable cause based on an anonymous letter that provided considerable predictive detail, but only after some of the detail was corroborated by police).

articulable reason to act).[20] In that scenario, police might still be prohibited from requisitioning the capacious digital record described above. But perhaps they would be justified in seeking more limited transactional data, say information about whether, near the predicted time, Slade calls a particular number or heads toward a particular location.

This measured approach to accessing The Cloud is based on what might be called the proportionality principle.[21] Under traditional Fourth Amendment rules, an arrest requires probable cause, but a short detention only reasonable suspicion; similarly, a full search of the person requires probable cause, a frisk only reasonable suspicion.[22] Analogously, significant invasions of privacy on The Cloud—obtaining a month's worth of bank records or internet logs, or as the Supreme Court itself suggested in *Jones*, travel records that track a person for four weeks[23]—might require cause about the target akin to that necessary to search a home or car; however, less significant invasions—accessing records about a single phone call, credit card purchase, or car trip, pulling up an identity using facial recognition technology, or tracking a car for a few hours—could be justifiable on something less. Not only does this type of proportionality principle better reflect the degree of the government's intrusion, but it also avoids the Catch-22 of requiring police to demonstrate probable cause before carrying out the preliminary investigative techniques they need to develop it.

Proportionality reasoning makes sense in the abstract. But it presents difficult line-drawing problems. What justification do police need if, rather than seeking data about Slade's financial transactions or travels over the course of a month, they want only a week's worth of data? Or if they want to ascertain, in combination, whether Slade calls a particular number, visits a particular location, and deposits a large amount of money during a given month, but seek no other information about him?

Answers to these types of questions inevitably produce somewhat arbitrary classifications. One approach is to differentiate between types of information. Perhaps medical records would receive the most protection, bank records something less, utility records something less still.[24] Current federal law appears to adopt this approach with respect to communications, with phone numbers and email addresses receiving minimal protection, subscriber information receiving more protection, stored communications even more, and interception of communications requiring probable cause.[25] But the intuition upon which this scheme is based is suspect: for instance, a months'-worth of "metadata" about a person's contacts may reveal much more than the transcript of a conversation.[26] Similar comments can be

[20] *See* Terry v. Ohio, 392 U.S. 1, 27 (1968).
[21] *See* Slobogin, *supra* note 4, ch. 2.
[22] *See Terry*, 392 U.S. at 20-27.
[23] *Jones*, 132 S.Ct. at 948.
[24] For an effort in this vein, see *Standards Governing Law Enforcement Access to Third Party Records*, A.B.A., http://www.americanbar.org/groups/criminal_justice/standards/law_enforcement_ access.html.
[25] *See supra* notes 2-4.
[26] Steven M. Bellovin, Matt Blaze, Susan Lanau & Stephanie K. Pell, *It's Too Complicated: How the Internet Upends Katz, Smith and Electronic Surveillance Law*, 30 HARV. J. L. & TECH. 1, 92 (2016) (given technological developments, "[t]he concept of metadata as a category of information that is wholly distinguishable from communications content and thus deserving of lower privacy protection is no longer tenable.").

made about other types of data: bank records, credit card statements, and utility logs can all be more or less private depending on the person and the context.

Sometimes the solution to this problem might be categorical. That was the angle the Supreme Court took with respect to searches of home interiors carried out with sophisticated technology; Justice Scalia, writing for the Court in *Kyllo v. United States*,[27] held that *all* such searches require probable cause. Government access to Cloud data that is analogous to the interior of the home—for instance, documents on one's computer that might also be stored on The Cloud; communications on a closed social network—should receive similar categorical protection.[28]

However, in other situations a supplemental approach to proportional regulation might rely on durational or aggregational limitations. In *Jones*, five members of the Court distinguished between "short-term" and "prolonged" tracking.[29] Similarly, the Court has indicated that, while a physical seizure lasting less than 15 minutes usually requires reasonable suspicion, a longer seizure amounts to an arrest requiring probable cause,[30] and an arrest must be judicially reviewed within 48 hours.[31] One might limit Cloud searches of non-public data outside the home context the same way, on the theory that the more one learns about a person—from whatever source—the more intrusion occurs. For instance, obtaining information about the transactions of someone like Slade on a particular day or over a couple of days might be permitted on a relevance showing, but seeking data shadowing his activities over more than a 48-hour period or with respect to several different days might require greater suspicion and a subpoena from a judge, and obtaining a months'-worth of transactions could require probable cause and a warrant. While this duration-based rule also has administrability problems,[32] it has the benefit of simultaneously protecting privacy in a roughly proportionate manner and permitting government to build its case without requiring probable cause from the outset.

Consider how this regime would work in connection with the National Security Agency's metadata program made famous by Edward Snowden, in particular the "two-hop" rule that now governs the program. Under the two-hop rule, once the NSA obtains a "seed identifier"—a phone number or electronic address associated with a national security threat—it is authorized to access the metadata of every connection with the identifier and, once those contacts are

---

[27] 533 U.S. 27, 37-38 (2001).

[28] Some have argued that encrypted material should receive *absolute* protection. But given the fact that anything, including impersonal business records, can be encrypted. proportionality reasoning would suggest that the government should be able to force decryption of any material for which it has the requisite cause. The encryption debate is too complicated to address in this limited space. *See* Hugh J. McCarthy, *Decoding the Decryption Debate: Why Legislating to Restrict Strong Encryption Will Not Resolve the "Going Dark" Problem*, 20 No. 3 J. INTERNET L. 1 (2016) (detailing practical problems and domestic and international legal issues associated with different approaches designed to permit government decryption).

[29] 132 S.Ct. at 964 (Alito, J., concurring).

[30] *See generally* United States v. Sharpe, 470 U.S. 675, 684-88 (1985).

[31] Cty. of Riverside v. McLaughlin, 500 U.S. 44, 56 (1991).

[32] *Compare* Orin Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311 (2011) (describing some of the problems), *with* Christopher Slobogin, *Making the Most of* United States v. Jones *in a Surveillance Society: A Statutory Implementation of Mosaic Theory*, 8 DUKE J. CONST.L L. & PUB. POL'Y 1, 24-30 (2012) (providing a model statute implementing mosaic theory).

identified, it is further authorized to go out one more hop to determine the numbers or addresses that have connected with the group in the first hop.[33] Under duration-based proportionality reasoning, if the NSA has probable cause, found by a judge, to believe the seed identifier is a number or email address that belongs to a terrorist, it might be justified in seeking many months' worth of that person's metadata; after all, the cause finding probably makes that person subject to arrest if authorities so chose. While this metadata also provides information about the suspected terrorist's contacts, that disclosure—the fact that on one or more occasions those persons communicated with the seed identifier—is minimal.

However, collecting all the contacts of everyone identified in the first hop, based solely on the initial finding of probable cause, is more problematic. The mere fact that the group that is discovered as a result of the first hop has been in contact with the seed identifier does not make them arrestable; at most it should only permit limited inquiries, because it stretches even the notion of reasonable suspicion to the breaking point. Certainly, a three-hop rule, which the NSA at one time purportedly followed, would be impermissible under proportionality reasoning.[34]

Even if one finds this type of reasoning persuasive in the abstract, it might be resisted in the specific context of NSA investigations. Where national security is at stake, the argument goes, any significant limitation on Cloud usage should be jettisoned. But this stance should be viewed with skepticism. "National security" is an extremely capacious term, and it has too often been a blank check for government abuse.[35] Concrete threats to the country might justify departure from the rules that normally govern domestic law enforcement; for instance, if there is a demonstrable, significant, and imminent danger associated with a seed identifier, relaxation of the justification required by proportionality reasoning might be permissible in this context.[36] But otherwise the NSA should probably be treated no differently than other law enforcement agencies.

II. Profile-Driven Cloud Access—Hit Rates

Profile-driven searches are very similar to suspect-driven searches. The difference is that suspect-driven searches start with a person thought to be engaged in wrongdoing and then go to The Cloud, while with profile-driven searches the government has no particular suspect when it seeks out Cloud-data; rather it utilizes a profile describing the characteristics of likely perpetrators that it hopes will identify wrongdoers. Again using John Slade as an example, imagine that the police focus on him not because of an anonymous tip but because of a drug dealer profile developed with the help of computer scientists and criminologists. Such a profile might be composed, let's say, of five factors having to do with travel, spending, and

[33] The Administration's Use of FISA Authorities: Hearing Before the H. Comm. on the Judiciary, 113th Cong. 36 (2013) (statement of John C. Inglis, Deputy Director, National Security Agency).
[34] By the third hop, over a million people might be affected. *Three Degrees of Separation: Breaking Down the NSA's 'Hops' Surveillance Method*, THE GUARDIAN (Oct. 28, 2013, 11:58 AM), http://www.theguardian.com/world/interactive/2013/oct/28/nsafilesdecoded-hops (showing that a person with fifty Facebook friends would link to 8170 second-tier people and 1,334,978 third-tier people).
[35] *See generally* Erik K. Yamamoto, *White (House) Lies: Why the Public Must Compel the Courts to Hold the President Accountable for National Security Abuses*, 68 L. & CONTEMP. PROBS. 285 (2005).
[36] I have called this the "danger" exception and sketched its parameters in Slobogin, *supra* note 4, at 26-28.

communication patterns. Or, similar to how credit card companies identify theft and fraud, the profile might purport to tell police when and where a drug deal is occurring or is soon likely to occur, which allows them to conduct surveillance of that spot and perhaps catch a perpetrator. Analogous to how researchers have developed risk assessment instruments for pretrial detention and sentencing purposes,[37] these profiles would initially be based on analysis of drug dealer characteristics and behavior, and then cross-validated on new populations or locations.

Such profiles are only useful, of course, if the government has access to databases that have the information needed to run the profile. Whether it should have such access is discussed below (under program-driven Cloud searches). Assume for now the data is available to government officials.

As with suspect-driven Cloud searches, the analysis of profile-driven Cloud inquiries should involve balancing what the police intend to do against the justification they have for doing it. If the police want to arrest anyone who fits the profile, the profile should give them probable cause to do so. If instead they want to identify people who will then become targets of further searching, either on The Cloud or through traditional means, the proportionality rule just discussed could apply. In either case, the question arises as to whether the profile can produce the necessary cause.

Usually, probable cause or reasonable suspicion is based on what the courts have called "individualized" suspicion, meant to connote the idea that the arrest, search, or stop in question is bottomed on facts particular to the individual arrested, searched, or temporarily detained. In a profile situation, in seeming contrast, the requisite cause is based on an algorithm developed through statistical analysis of criminals or crimes, in an effort to develop correlations that are predictive. The fact that profiles are derived from studies of groups and then used to identify a single individual, like Slade, who is not in the group has bothered some who believe profile-driven searches or seizures are not based on truly individualized cause and that this lack of individualization makes profile-driven searches illegitimate.[38]

Yet at bottom the only difference between the type of "generalized" or "categorical" suspicion incorporated into profiles and the type of suspicion associated with "individualized" determinations is that the former method of developing cause is more quantitative in nature. A person targeted because of a profile is still being targeted because of characteristics or behavior personal to him or her. At the same time, so-called "individualized" searches are, like profiles, inevitably based on stereotypes (another word for profiles). The classic stop and frisk—famously illustrated by the facts of *Terry v. Ohio*,[39] where an officer observed three men engaging in behavior consistent with casing a store for a robbery—may be based on "particularized" suspicion, as the Supreme Court later put it.[40] But such stops are also based on common

---

[37] *See e.g.*, Christopher T. Lowenkamp & Jay Wetzel, *The Development of an Actuarial Risk Assessment Instrument for U.S. Pretrial Services*, 73-SEP FED. PROBATION 33 (2009).
[38] *See* Kiel Brennan-Marquez, *Plausible Cause*, 70 VAND. L. REV. ___ (forthcoming 2017) (arguing against profile-based searches and seizures).
[39] 392 U.S. 1 (1968).
[40] I.N.S. v. Delgado, 466 U.S. 210, 233 (1984).

knowledge—in *Terry* itself, the officer's knowledge about how robbers case a store. As Fred Schauer pointed out, "[O]nce we understand that most of the ordinary differences between general and particular decisionmaking are differences of degree and not differences in kind, we become properly skeptical of a widespread but mistaken view that the particular has some sort of natural epistemological or moral primacy over the general."[41]

Profiling using data accumulated from Cloud-related sources, sometimes called "predictive policing," is in its infancy. But police departments appear to be committed to developing the necessary tools.[42] As that development continues, it should be limited in at least three ways.[43] First, as already suggested, the profile must produce a "hit rate" equivalent to the certainty required by the proportionality principle. A profile that correctly identifies a drug dealer only 20% of the time does not provide the probable cause necessary for search of a home. However, it might permit a suspect-driven Cloud search that stays under the two-day rule or is limited in some other way. And a profile that is right 50% of the time may provide grounds for a traditional search, at least if we are willing to associate probable cause with a more-likely-than-not standard (which we may not be, once it is quantified in this way).

This hit rate limitation could be a significant one. Achieving a 50% hit rate or even a 20% rate may be impossible for most crime scenarios; certainly social scientists engaged in the analogous pursuit of predicting dangerousness for sentencing purposes have struggled to achieve such accuracy.[44] There are scores of variables associated with criminal behavior, and the explanatory power of any given variable or combination of variables is likely to be very low. Further, profiles will probably need to be updated routinely, either because of naturally occurring changes in criminal behavior or because perpetrators get wind of the factors in the profile. When one adds to those challenges the fact that much of the information about individuals found on The Cloud is unreliable,[45] profiles that might justify apprehending specific suspects will be few and far between, at least if police action based on such data abides by the proportionality principle.

Assuming that profiles with acceptable hit rates can nonetheless be developed, a second limitation on profile-driven Cloud use is that it should be transparent. To avoid profiles concocted ex post, and to ensure that those individuals who are targeted using a profile actually meet it, profiles must be accessible to courts and other oversight entities, at least on an in camera basis. Transparency also assures that the factors on which profilers rely are vetted to ensure that illegitimate ones, such as race, are not influencing the results.

---

[41] FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 69 (2003).
[42] Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 352-88 (2015).
[43] Two other limitations, both applicable to all of The Cloud searches discussed here and both crucial to assuring accountability, is an auditing program that keeps a record of who accesses data when and for what purpose and a notice requirement that would alert those whose data have been accessed, ex post if not ex ante.
[44] *See* Douglas Mossman, *Assessing Predictions of Violence: Being Accurate About Accuracy*, 62 J. CONSULTING & CLINICAL PSYCHOL. 783 (1994).
[45] Ferguson, *supra* note 42, at 398-99.

This vetting process could become difficult if, as occurs in some commercial contexts, profiles rely on complex algorithms generated through opaque machine-learning techniques or protected from disclosure for proprietary reasons.[46] Complicating matters further, risk factors such as criminal history, location, and employment may turn out to be proxies for race, class, and related traits, use of which are generally considered anathema in police work.[47]

These concerns do not have to be paralyzing, however. For instance, profiles that are indecipherable could be banned in the law enforcement context, regardless of their accuracy,[48] or can be designed to ensure "procedural regularity."[49] Steps can also be taken to alleviate the concern that some risk factors correlate with race as well as crime.[50] For instance, developers of algorithms designed to detect potential perpetrators or crime hot spots could be directed to avoid arrest records for low-level or drug crimes that might reflect race-based policing practices; instead, developers can be told to rely on *reports* of crimes (for hot spot profiles) and on crimes of violence or on property crimes (for suspect profiles), so as to reduce the influence of racially-biased arrest rates for drug crimes and similarly manipulable offenses.[51] It is also important to remember that traditional policing often relies on the same suspect, static factors, in ways that are inevitably more intuitive, and therefore less discoverable and more subject to invidious manipulation. Transparent algorithms that can produce the relevant hit rates and that avoid obviously illegitimate variables are very likely to be an improvement.[52]

To limit further the extent to which bias creeps into the process, however, a third limitation that should be imposed on profile-driven Cloud searches is the maxim that everyone who fits a given profile must be treated in the same neutral fashion. That means if a drug dealer profile with the relevant hit rate identifies 200 people, police should not be able simply to single out someone like Slade but rather would either have to investigate everyone who fits the profile, or if that is not feasible, select individuals on a random basis (e.g., every third person). In the

---

[46] *See* Michael L. Rich, *Machine Learning, Automated Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 883-86 (2016).
[47] Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).
[48] The science of "interpretable machine learning," designed to ensure that algorithms are understandable to humans, is already being applied in medical settings. Been Kim, *Interactive and Interpretable Machine Learning Models for Human Machine Collaboration*, http://people.csail.mit.edu/beenkim/papers/BKimPhDThesis. pdf; MICROSOFT, *2016 Workshop on Human Interpretability in Machine Learning*, https://sites.google.com/site/ 2016whi/.
[49] Joshua A. Kroll, et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017) (sketching how computer programs can be constructed to ensure fairness and procedural regularity despite the black box nature of machine learning).
[50] See the debate on this issue summarized in Anthony W. Flores, Christopher T. Lowencamp & Karen Bechtel, *False Positives, False Negatives and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals, and It's Biased against Blacks*" (2016), http://www.crj.org/page/-/publications/rejoinder7.11.pdf.
[51] *Cf.* Michael Feldman, et al., *Certifying and Removing Disparate Impact,* KDD '15 Proceedings of the 21th International Conference on Knowledge Discovery and Data Mining 259-268 (2015) (discussing similar moves in connection with hiring algorithms).
[52] *See, e.g.*, Sharad Goel, Maya Perelman, Ravi Shroff & David Alan Sklansky, *Combatting Police Discrimination in the Age of Big Data*, https://5harad.com/papers/policing-the-police.pdf (using stop and frisk data from New York City to create a risk profile that vastly reduced the number of seizures while producing a much higher hit rate; also finding that factors like "furtive movement," a common police justification for stops, was not related to weapon possession and that, of those stopped using the profile, whites were much more likely than blacks to have a weapon).

absence of this limitation, attempts to avoid illegitimate discrimination in construction of the profile will merely reappear at the post-profile investigation stage.

The added advantage of this third limitation on profile-driven actions is that it would make law enforcement think twice before engaging in them. Profile-driven searches will produce a large number of false positives, no matter how good they are. If, for instance, the predicted hit rate is 50%, half of those investigated are likely to be innocent, whether the police go after everyone identified by the algorithm or only a randomly selected subgroup. Even if the post-profile police work is covert, much investigative energy will be expended with no gain. And in those situations where the investigation of those who meet the profile involves overt searching or seizing, a non-trivial number of false positives are likely to complain. Although the quantified, objective nature of profile-driven Cloud searches offers many advantages over traditional suspect-based techniques, their dragnet nature may end up being so practically or politically unpalatable that police abandon them.[53]

III. Event-Driven Cloud Access—Hassle Rates

Some Cloud searches conducted by law enforcement start not with a suspect or a profile of a likely suspect, but with an event—usually a crime—and use Cloud data to try to figure out who perpetrated or witnessed it. Let's return to the example of John Slade, but this time as a victim rather than a potential suspect. Imagine that, at 2:00 a.m. one Sunday morning, police are called to the scene of a homicide, a dark urban street, where they find Slade dead, drugs strewn around him. A medical examiner says the death probably occurred two hours earlier, around midnight. Pre-Cloud, the police would probably go door to door talking to those who live in the immediate vicinity, some or all of whom might claim—honestly or not—to have been elsewhere at the relevant time or to have seen or heard nothing. In contrast, today police might access phone or vehicle GPS records, as well as feeds from closed-circuit TV or airborne cameras with face-recognition or night vision capacity, to identify people or cars near the crime scene at the time it happened, and then use suspect-driven techniques to zero in on the perpetrator.[54]

These event-driven uses of The Cloud could result in a large haul of people, among whom may be the perpetrator or a witness, but many of whom will be neither. At the same time, all that this "data dump" learns about any of these individuals is that they were near a particular place at a particular time, a discovery that proportionality reasoning would suggest requires little justification. Even so, the government's Cloud net should probably be limited, to reduce both the extent of the initial privacy invasion and the number of people subject to further law enforcement inquiry. In other words, the government should minimize what Jane Bambauer calls the "hassle

[53] *See* William Stuntz, *Implicit Bargains, Government Power, and the Fourth Amendment*, 44 STAN. L. REV. 553, 588 (1992) (applying a "politics model" of the Fourth Amendment to search and seizure scenarios involving large groups of people that relies on the electorate to "throw the rascals out" when the program becomes onerous).
[54] Baltimore has used videos from plane cameras to "TiVo" backward from the scene of the crime to determine how individuals and vehicles got there. *See* https://www.bloomberg. com/features/2016-baltimore-secret-surveillance/ (last accessed, Oct. 1, 2016).

rate"—the proportion of innocent people subject to police investigation in an effort to find the one or two bad people.[55]

What that rate should be will depend on the likely number of people involved. In effect, an admonition to limit hassle rates is simply a call to shape event-driven searches around the relevant time and place. In investigating Slade's death, for instance, police should be able to find out the identity of and question pedestrians and car drivers near the scene of the crime shortly before or after midnight (if the medical examiner's assessment is correct) but perhaps should not be able to investigate people who never approached the scene closer than 50 yards or who were there before 11:30 p.m. or after 12:30 a.m.

Other event-driven searches might call for more difficult decisions about hassle rates. For instance, assume several people are killed by a sniper. Ballistics indicates that the weapon used is relatively rare, and police also deduce from the way the killings took place that the sniper has a very powerful scope. In an effort to discover potential suspects, may law enforcement demand from every gun store in the vicinity files on the identities of everyone who has bought that type of gun or scope? Or assume police know that a bomb explosion occurred in a particular type of van. May police discover the identities of everyone who owned or rented such a vehicle within 500 miles of the explosion? Or, to take an actual case, may police search residential records of all males who lived in both Philadelphia, Pennsylvania, and Fort Collins, Colorado, during the time periods that several sexual assaults with the same modus operandi occurred in those two cities?[56] In all of these cases, as in the Slade example, investigators find out only tidbits of information about any given individual. But the tidbits all involve information that at least some of those people are likely to want to keep private from government snooping even if they are innocent. More importantly, subsequent face-to-face investigation of those who are discovered this way will involve very high hassle rates.

The Cloud facilitates immensely the ability of investigators to carry out event-driven inquiries. As these examples illustrate, such inquiries can be quite broad, limited only by the imagination and priorities of law enforcement (because they are not limited by current law, at least in most jurisdictions). In contrast to the hit rates required for profile-driven Cloud searches, acceptable hassle rates for event-driven Cloud searches are not easy to establish, and should probably vary with the type of information sought and the type of crime being investigated.[57] If the law is called into play here, perhaps the best that can be done is to require police to seek authorization for such inquiries from a judge, who can take potential hassle rates and these other factors into account in determining whether and to what extent event-driven Cloud searches may occur.

---

[55] Jane Bambauer, *Hassle*, 116 MICH. L. REV. 461 (2015).

[56] *See* William J. Krouse, *The Multi-state Anti-terrorism Information Exchange (MATRIX Pilot Project)*, Congressional Research Service Report RL325369 9 (Aug. 18, 2004).

[57] In an analogous situation, the Supreme Court held that the analysis should consider "the gravity of the public concerns served by the seizure, the degree to which the seizure advances the public interest, and the severity of the interference with individual liberty." Illinois v. Lidster, 540 U.S. 419, 427 (2004) (upholding a roadblock at the time of day and the place of a hit-and-run accident committed one week earlier, set up to find possible witnesses).

IV. Program-Driven Cloud Access—Democratic Authorization

Suspect-driven, profile-driven, and event-driven Cloud searches all rely in varying degrees on access to multiple databases, ranging from those that keep track of communications and travels to those that house records of financial and social transactions. From law enforcement's perspective, keeping these databases within their separate silos is, at the least, inefficient and, in the case of profile-driven Cloud access, perhaps fatal, since profiles usually only work when they can access several databases at once. It was in recognition of this fact that the Defense Department proposed, post-9/11, the Total Information Awareness (TIA) program. According to a chart prepared by the Department of Defense, TIA was meant to gather in one place a huge amount of transactional data concerning, according to the official description, information about "financial, educational, medical, veterinary[!], entry [i.e., immigration and customs], transportation, housing, . . . and communications" activities, as well as all government records.[58] Once collected, these data would be combed using algorithms designed to detect terrorist activity. Congress, apparently not enamored of this idea, defunded TIA in 2003 (by voice vote).[59] But if Edward Snowden is to be believed, several programs in operation today, run by the NSA or other government agencies, bear at least some resemblance to it.[60]

As the public reaction to Snowden's revelations indicates, a significant proportion of the citizenry is uncomfortable with these types of programs. Compilation of information from multiple sources in one "place" raises a host of concerns. As recent exposés of foreign machinations highlight, aggregation of data facilitates hacking and identity theft.[61] It also leads to "mission creep," as law enforcement realizes that information obtained for one reason (such as fighting terrorism) might be useful to other purposes. It can easily lead to more obvious abuses, ranging from illegitimate investigations of journalists, politicians, activists, and members of certain ethnic groups to leaks based on personal vendettas.[62] Most prominently, it tempts the government to combine all of the information it has collected to create "personality mosaics" or "digital dossiers" about each of its citizens, a phenomenon classically associated with totalitarian states.[63]

---

[58] *See* https://www.google.images (prompt: Total Information Awareness).
[59] *See* 149 Cong Rec S 1379-02, 1416 (Jan 23, 2003).
[60] *See* Glenn Greenwald, *XKeyscore: NSA Tool Collects "Nearly Everything a User Does on the Internet,"* THE GUARDIAN, July 31, 2013, http://www.theguardian.com/world/2013/jul/31/nsa-to-psecret-program-online-data.
[61] *See, e.g.*, Nicole Perlroth & David Gelles, *Russian Hackers Amass over a Billion Internet Passwords*, N.Y. TIMES (Aug. 4, 2014).
[62] For some examples, *see* Robert H. Sloan & Richard Warner, *The Self, the Stasi, and the NSA: Privacy, Knowledge, and Complicity in the Surveillance State*, 17 MINN. J. LAW SCI. TECHNOL. 347-380 (2016) ("The government uses the information it has to discourage and prevent behavior of which it disapproves" including behavior by "journalists, political dissenters, lawyers representing political activists and dissenters, politicians opposing the policies and goals of those with the power to order surveillance, sustainable energy advocates, environmentalists, animal rights activists, Afro-Americans, Muslims, labor unions, people seeking health care, welfare recipients, parolees, and a diverse collection of types of people the government regards as (possibly) undesirable") (citations to examples omitted).
[63] Daniel Solove popularized the term "digital dossiers," which he described as the aggregation of data to create "a profile of an individual's finances, health, psychology, beliefs, politics, interests, and lifestyle" that "increasingly flows from the private sector to the government, particularly for law enforcement use." Daniel J. Solove, *Digital Dossiers and the Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1084 (2004).

In part because of the public reaction to Snowden's disclosures, the NSA has supposedly divested itself of the metadata it had been collecting and must now seek it through subpoenas from the relevant common carriers, in the suspect- and profile-driven manner described earlier.[64] But the NSA and other federal agencies continue to aggregate other types of data.[65] Localities and states also engage in the data collection enterprise. For instance, New York City's Domain Awareness system, co-created by the city's police department and Microsoft, collates information gleaned from thousands of closed-circuit surveillance cameras (CCTV), and combines it with geospatial data that reveals crime "hot spots," feeds from license recognition systems, and GPS signals that permit real-time and historical tracking of cars.[66] A number of other cities operate large-scale CCTV systems, and many are also moving toward 24/7 drone or plane surveillance.[67] A different type of program, known as the "fusion center," exists in more than half the states. These centers—over 75 at last count, some with more than 200 personnel—"fuse" financial, rental, utility, vehicular, and communications data from federal, state, and local public databases, law enforcement files, and private company records for investigative purposes.[68]

These program-driven efforts, which have been called "panvasive" because they invade the records of large swaths of the population, occur with the foreknowledge that most of those affected have done nothing wrong.[69] Thus, this collection of data cannot be regulated through suspicion-based proportionality reasoning. Arguably, however, it does not need to be. Until the data is accessed by humans and used as a means of investigating or identifying particular people like Slade, no concrete intrusion has occurred. Only when such access does occur need government officials demonstrate the cause necessary to carry out suspect-, profile-, or event-driven searches.

For those who do not trust government to abide by such strictures, one further protection, illustrated by Congress' changes to the NSA's metadata program, would be to require that all databases be maintained outside the government. Even profile-driven Cloud searches could be carried out by a private entity, with the government providing the profile and the company providing the government only with the identities of those who meet it. While this arrangement would still present some of the problems associated with aggregation (hacking and the like), it would undoubtedly reduce the potential for mischief by government officials.

---

[64] USA Freedom Act, Public Law 114-23, § 101 (June 2, 2015).
[65] Zack Whittaker, *Freedom Act Will Kill Only One of NSA's Programs (and Not Even One of Its Worst)*, ZERO DAY (May 4, 2014) http://www.zdnet. com/article/ freedom-act-metadata-phone-records-prism/#!
[66] *See* Colleen Long, *NYPD, Microsoft Create Crime-Fighting "Domain Awareness" Tech System*, HUFFINGTON POST (Feb. 25, 2013).
[67] On the increase in CCTV, *see* Semini Sengupta, *Privacy Fears Grow as Cities Increase Surveillance*, N.Y. TIMES, Oct. 13, 2013. On the increase in drone surveillance and attempts to regulate it, *see* Marc Jonathan Blitz, James Grimsley, Stephen A. Henderson & Joseph Thai, *Regulating Drones under the First and Fourth Amendments*, 57 WM. & MARY L. REV. 49 (2015).
[68] *See* THE CONST. PROJECT, RECOMMENDATIONS FOR FUSION CENTERS: PRESERVING PRIVACY AND CIVIL LIBERTIES WHILE PROTECTING AGAINST CRIME AND TERRORISM 4 (2012), www.constitutionproject.org/pdf/fusioncenterreport.pdf (describing the establishment of 77 fusion centers nationwide and the types of information these centers collect).
[69] *See* Christopher Slobogin, *Panvasive Surveillance, Political Process Theory and the Nondelegation Doctrine*, 102 GEO. L. J. 1721 (2014).

In the end, however, this attempt to separate government from data cannot work. Many of the databases useful to Cloud searches—those that house CCTV feeds, the data from highway tracking systems, and the billions of personal records relevant to criminal history, taxes, entitlements, real estate transactions, and scores of other matters—would not exist but for the government. The executive branch needs this information for all sorts of legitimate reasons, some related to crime prevention and many that are not. Government should not be prohibited from collecting and maintaining it.

Instead, regulation of program-driven Cloud searches must come from the political process.[70] Given Congress' supineness to executive branch surveillance proposals after 9/11, that suggestion may seem naïve. But legislatures are capable of action in this area, as the defunding of TIA and the revamping of the NSA's metadata program illustrate.[71] Especially when, as is the case with many types of Cloud-based efforts, the program affects significant segments of the population—including members of the legislature and their most powerful constituents—some type of political oversight is not only possible but likely.

At the same time, it must be admitted that law enforcement and tough-on-crime lobbies are a forceful presence at both the federal and state levels and may be able exert influence that the populace as a whole cannot. That is where the courts could come into play, in two ways. On rare occasions, courts might declare a particular data collection scheme unconstitutional under the Fourth Amendment. However, given the Supreme Court's narrow definition of the word "search" for Fourth Amendment purposes and its high level of deference even to programs that it is willing to say involve searches (under what it calls its "special needs" jurisprudence[72]), that outcome is not likely in the near future. A second way courts might nudge legislatures and law enforcement agencies toward a balanced view—and one that would operate independently of the Fourth Amendment—is by applying the same "hard look" analysis they apply to programs created by other administrative agencies.[73] While law enforcement departments have seldom been subject to the type of judicial monitoring to which other agencies routinely submit, that lack of oversight is likely an historical accident rather than a considered policy. The full argument for why courts are obligated to engage in such oversight will not be set out here.[74] For present purposes, it suffices to say that, where program-driven, panvasive operations are involved, a solid case can be made that the courts should treat police agencies the same way they treat other agencies that are engaged in creating rules governing the circumstances under which people may carry out innocent conduct.

---

[70] *Id*. at 1745-58.

[71] Other examples come from state statutes that limit the use of drone surveillance, Michael L. Smith, *Regulating Law Enforcement's Use of Drones*, 52 HARV. J. LEGIS. 423, 427-432 (2015) (cataloguing state drone statutes), and federal statutes that require subpoenas for a number of different records. Erin Murphy, *The Politics of Privacy in the Criminal Justice System: Information Disclosure, the Fourth Amendment, and Statutory Law Enforcement Exemptions*, 111 MICH. L. REV. 485 (2013).

[72] For a description of this jurisprudence *see* Slobogin, *supra* note 69, at 1727-33.

[73] See, *e.g.*, Patrick M. Garry, *Judicial Review and the "Hard Look" Doctrine*, 7 NEV. L.J. 151, 154-59 (2006) (discussing the Administrative Procedure Act and associated case law establishing the hard look doctrine).

[74] *See* Christopher Slobogin, *Policing as Administration*, 165 U. PA. L. REV. 91 (2016).

That conclusion has several consequences. First, under accepted administrative law principles, no agency program that affects the rights and obligations of the citizenry may exist unless the agency can point to authorizing legislation that sets out the harm to be prevented, the persons and activities likely to be affected, and the general means for preventing the harm. That would mean that before programs like New York City's Domain Awareness operation and the states' fusion centers can come into being, municipal, state, or federal legislatures would have to think through the types of information they can obtain and for what purpose. That requirement of legislative authorization, enforced by the courts, would ensure at least some democratic assessment of such programs and how they should operate.

The impact of administrative law principles would not end there, however. Standard practice dictates that, once authorized to set up a program, an agency must draft implementing rules, subject them to a notice and comment process (or something similar) that allows public input, and provide written rationales for the rules ultimately chosen, rules that are reviewable by a court to ensure they meet a demonstrated need and that they are applied even-handedly, without irrational distinctions between groups or areas.[75] This further injection of democratic input and judicial oversight would exert significantly more pressure on police departments to consider competing views when contemplating the creation of a data collection scheme. Regulated through this type of public process, it is likely that TIA-like programs, fusion centers, and other panvasive practices would be significantly curtailed or implemented with more care.

One drawback to the political process approach to program-driven Cloud searches is the possibility its transparent nature will enable the bad guys to learn the ins-and-outs of the programs and how to avoid them. But this traditional law enforcement concern, which administrative procedure acts specifically recognize as legitimate,[76] is exaggerated in this setting. The primary aim of most panvasive actions is deterrence, which publicity can only enhance. Further, matters of specific implementation need not be revealed. For instance, if camera surveillance is meant to be covert, the fact and general area of such surveillance should be disclosed, but exact camera locations need not be. The types of records sought by fusion centers should be revealed, but the algorithms that might be used to analyze them could be viewed in camera. Ultimately, however, the primary response to the tip-off concern is that democratic accountability requires that the public be told not only what panvasive capacities police have but how those capacities will be used.

V. Volunteer-Driven Cloud Searches—Fiduciary Obligations

All of the foregoing Cloud searches involve government-initiated investigations. The assumption throughout this paper has been that when the government decides to intrude, some justification is necessary. But what if a data-holder—a bank, a common carrier, or hospital—comes across information it thinks is indicative of criminal activity and wants to hand it over to the police? While the discussion thus far has suggested several reasons why government should

[75] *Id.* at 45-57.
[76] *See, e.g.*, 5 U.S.C. § 552(b)(7)(E) (2012).

not be able to demand information from a third party without justification, the situation is clearly different when the third party comes forward of its own accord.

Even so, it is important to recognize that not all volunteer-driven Cloud searches are alike. In the cases in which the Supreme Court first announced the third party doctrine, the third party was a personal acquaintance of the defendant.[77] Establishing a rule that the government must ignore disclosures from such people denigrates their autonomous choice to make the disclosures, and could even be said to undermine their First Amendment right to speech. Recall, for instance, the tipster in the first hypothetical involving John Slade. Whatever that person's motives and however that person acquired the information, the choice to divulge it deserves respect and should be considered a legitimate basis for government action if it has sufficient indicia of reliability.

However, in the Court's later third party cases, *Miller v. United States*[78] and *Smith v. Maryland*,[79] the third party was not a person but an institution, more specifically, a bank and a phone company. Historically, corporations have not been considered autonomous "persons" in most contexts and have also been accorded lesser First Amendment rights than natural beings.[80] More importantly, unlike human confidantes, these institutions can be said to owe either formal or quasi-formal fiduciary duties to their customers, because unlike the human third party, they are able to obtain personal facts solely because they purport to provide a particular service.[81] The most sympathetic example on point comes from the medical context, where a patient provides information to a treatment provider. Even the Supreme Court has balked at the notion that a hospital is entitled to ignore a patient's expectation of medical privacy for the purpose of nabbing criminals.[82] Arguably, an analogous position is warranted with respect to banks and phone companies, to which we give information for the sole purpose of carrying out financial transactions or communicating.

Also important to recognize is that, when the third party is an institution, the degree to which information is "voluntarily" handed over to the government can vary greatly. In some cases, the government *commands* third parties to produce information about others, automatically and in the absence of a particularized court order. For instance, banks must report all deposits of $10,000, regardless of circumstances.[83] If this sort of dictate is justifiable, it should be so only if it comes from the legislature and is generally applicable (as is true in the deposit scenario). More commonly, the government exerts subtler pressures on third parties to produce information. Most obviously, some data brokers, although purportedly private and

[77] *See, e.g.*, Lewis v. United States, 385 U.S. 206 (1966); Hoffa v. United States, 385 U.S. 293 (1966).
[78] 425 U.S. 435 (1976).
[79] 442 U.S. 735 (1979).
[80] Citizens United v. Fed. Elec. Comm'n, 558 U.S. 310 (2010) focused on *political* speech rights of corporations, which are not implicated in this context. Further, corporations are still not considered "persons" for Fifth Amendment purposes, Hale v. Henkel, 210 U.S. 43 (1906), and have very weak Fourth Amendment rights. United States v. Morton Salt Co., 338 U.S. 632, 652 (1950).
[81] *See* Kiel Brennan-Marquez, *Fourth Amendment Fiduciaries*, 84 FORDHAM L. REV. 61 (2015).
[82] Ferguson v. City of Charleston, 532 U.S. 67 (2001).
[83] 31 U.S.C. § 5313(a).

independent of the government, essentially see the government as their client,[84] and other companies, dependent on government largesse, may be especially eager to show they are helpful.[85] Unless defined narrowly, volunteer-driven Cloud searches might ultimately even undo efforts, like the recent NSA legislation, to keep as much data as possible out of government hands. That phenomenon is worrisome, because people should be able to trust that the private institutions on which they depend for the basics of life are not conduits to the government.

At the same time, it must be acknowledged that fiduciary obligations and concerns about corporate duplicity should not always trump speech rights and concerns about public safety. For instance, both the medical and legal professions recognize a duty to reveal information that would prevent a violent crime or forestall an ongoing one.[86] Explicitly applied to The Cloud, that norm would permit third party institutions to disclose, and government to use, information about others that is likely, if known by the government, to prevent a serious violent felony from taking place in the near future. Arguably, however, that norm should be the full extent to which the law bows to the volunteer notion where third party institutions that are essential to living in the modern world are involved (a position recognized in at least one federal statute).[87]

VI. Another Route: Regulating Disclosure Rather than Access

An alternative to all of this, perhaps initially attractive, is to let the government obtain and peruse all the data it wants, without limitation, but severely curb what could be revealed to the public. For instance, shortly after 9/11 William Stuntz proposed that the government be permitted to engage in suspicionless data collection and access, but that it be prohibited from using information thereby discovered except to prosecute or detain terrorists and others who have committed violent felonies.[88] Since such data collection would take place covertly, only those suspected of committing serious crimes would have personal matters exposed to the public. In this type of Cloud-access regime, Stuntz argued, the benefits to security would outweigh any societal harms.

Under the rules proposed above, the data *collection* aspect of Stuntz's proposal might be permissible, if legislatively authorized and transparently regulated under administrative law principles. But untrammeled *access* to the data so collected would not be. In proposing otherwise, Stuntz seriously underestimated the harms of a regime that limits only use and not access. First, the citizenry would eventually figure out not only that government has data about

---

[84] Hoofnagle, *supra* note 5, at 617-18.
[85] Avidan Y. Cover, *Corporate Avatars and the Erosion of the Fourth Amendment*, 100 IOWA L. REV. 1441, 1445 (2015) ("technology corporations are not likely to challenge government surveillance requests, and even less likely to make effective arguments asserting their individual customers' rights, because of their government connections, the legal constraints on transparency and disclosure, and their immunity for complying with the government.").
[86] MODEL RULES OF PROF'L CONDUCT r. 1.6(b)(1) ("a lawyer may reveal information relating to the representation of a client to the extent the lawyer reasonably believes necessary . . . to prevent reasonably certain death or substantial bodily harm."); Fl. Stat. §394.4615(3)(a) ("[w]hen a patient has declared an intention to harm other persons," the therapist may release "sufficient information to provide adequate warning to the person threatened.").
[87] *See, e.g.*, 18 U.S.C. § 2702(c) (restricting the ability of private ISPs to disclose communications to law enforcement voluntarily to emergencies involving death or serious physical injury and a few technical situations).
[88] William J. Stuntz, *Local Policing After the Terror*, 111 YALE L.J. 2137, 2183-85 (2002).

everything it does, but that its agents are poring through that data, knowledge that is likely to chill innocent behavior and create a sense of oppression that is inimical to a free society.

Second, mission creep would be inevitable; law enforcement officials are not likely to ignore evidence of crime just because it falls short of being "serious," however that word is defined (and note that here, as opposed to the above proposal concerning volunteer-driven searches, use of Cloud data would not be limited to dealing with ongoing or future crime). If history is any guide, misuse of information for non-investigatory purposes would not be far beyond. Third, the conceit that only a small coterie of officials will be privy to the information collected in such a regime is naïve; post-9/11 there are literally thousands of officers authorized to handle anti-terrorist data which, if expanded to include all data, would vastly increase the potential for illegitimate leaks and abuse. Movement in the direction of a know-it-all government may be inevitable, but without currently unattainable advances in data accuracy, analytics, de-biasing techniques, and security, that movement should not be explicitly endorsed.[89]

Conclusion

Private and government databases are full of information that can enhance law enforcement's ability to detect and investigate crime and terrorism. Given the personal nature of much of this information, the police should not be able to obtain, view, or use it at will. If government wants non-public records about an identified suspect it should have to demonstrate it has suspicion proportionate to the intrusion involved. If it deploys a profile to identify suspects through a data search, it should ensure the profile produces the requisite proportionality-derived hit rate, avoids illegitimate discrimination, is open to inspection, and is used in the knowledge that everyone identified should be subject to further investigation. If government instead is relying on a crime as its starting point, its use of The Cloud should be attentive to hassle rates, keeping the number of people investigated to the minimum dictated by the time and place of the crime. Collections of data should be maintained outside of government to the extent consistent with governing needs, but wherever maintained should be authorized by specific legislation and administrative rules transparently and democratically arrived at and even-handedly applied. To discourage the government from creating financial or other incentives that seek to evade these limitations, private institutions should be permitted to proffer the government information about those to whom they own a de facto fiduciary duty only where it would prevent an ongoing or future serious felony. These rules allow the government to take advantage of The Cloud's investigative potential while cabining the temptation to abuse it.

[89] For an interesting, largely positive take on a surveillance state given the assumption that these goals are attainable, see Richard M. Re, *Imagining Perfect Surveillance*, 64 UCLA L. REV. DISCOURSE 264 (2016).