Metadata in Context –
An Ontological and Normative Analysis of the NSA’s Bulk Telephony Metadata Collection Program

Paula H. Kift
PhD Candidate
Media, Culture, and Communication
New York University
paula.kift@nyu.edu

Helen F. Nissenbaum
Director, Information Law Institute
Professor, Media, Culture, and Communication
New York University
helen.nissenbaum@nyu.edu

Abstract: In the aftermath of the Snowden revelations, the National Security Agency (NSA) responded to fears about warrantless domestic surveillance programs by emphasizing that it was collecting only the metadata, and not the content, of communications. When justifying its activities, the NSA offered the following rationale: because data involves content and metadata does not, a reasonable expectation of privacy extends only to the former but not the latter. Our paper questions the soundness of this argument. More specifically, we argue that privacy is defined not only by the types of information at hand, but also by the context in which the information is collected. This context has changed dramatically. Defining privacy as contextual integrity we are able, in the first place, to explain why the bulk telephony metadata collection program violated expectations of privacy and, in the second, to evaluate whether the benefits to national security provided by the program can be justified in light of the program's material costs, on the one hand, and its infringements on civil liberties, on the other hand.

Table of Contents

TABLE OF CONTENTS 1
I. INTRODUCTION 2
II. ONTOLOGICAL ANALYSIS 4
A. Defining Metadata 4
B. Classifying Metadata in the Law 6
   Content v. non-content 6
   Private records v. Business records held by third parties 9
   Hard-to-obtain information v. Information in plain view 10
III. NORMATIVE ANALYSIS 16
   A. Contextual Integrity 16
      Voluntariness 17
      Capabilities 22
      Assumption of risk 26
   B. Evaluation 29
   C. Outlook 31
IV. CONCLUSION 32
I. INTRODUCTION

A man writes a dozen letters to different people. No person would be permitted to publish a list of the letters written.


In the aftermath of the Snowden revelations, the National Security Agency (NSA) responded to fears about warrantless domestic surveillance programs by emphasizing that it was collecting only the metadata, and not the content, of communications. When justifying their activities, the NSA offered the following rationale: because data involves content and metadata does not, a reasonable expectation of privacy extends only to the former but not the latter. In other words, the NSA drew a normative conclusion about differential treatment of data and metadata based on an ontological distinction it claims exists between the two. Our paper questions the soundness of this argument. More specifically, we argue that privacy is defined not only by the types of information at hand, but also by the context in which the information is collected. This context has changed dramatically. Defining privacy as contextual integrity we are able, in the first place, to explain why the bulk telephony metadata collection program violated expectations of privacy and, in the second, to evaluate whether the benefits to national security provided by the program can be justified in light of the program’s material costs, on the one hand, and its infringements on civil liberties, on the other hand.

The first part of our paper traces the roots of the data/metadata distinction to the library and computer sciences, where metadata is characterized as data used to describe other data. In the aftermath of the Snowden revelations, by contrast, the courts struggled to characterize metadata in light of precedent. As a result, an assessment of whether or not the bulk collection of telephony metadata violates a reasonable expectation of privacy was rooted in three constitutionally relevant dichotomies, namely content vs. non-content data, private records vs. business records held by third parties, and hard-to-obtain information vs. information “in plain view.” Our paper traces the genealogy of cases that have influenced these distinctions in order to explain why the judges presiding over the two cases that have challenged the NSA’s program thus far – Judge Leon in *Klayman v. Obama* (2013) and Judge Pauley in *ACLU v. Clapper* (2013) – reached opposite conclusions as to whether the bulk telephony metadata collection program violates the Fourth Amendment. Our paper ultimately supports the argument of the *Klayman* court that even if the nature of metadata has not changed (and this is debatable), the circumstances in which
metadata is collected have. Whether or not the bulk collection of telephony metadata violates a reasonable expectation of privacy thus requires not only an *ontological* analysis of what metadata “is” but also an assessment of its *normative* significance in light of an evolving social and technological environment.

The second part of the paper engages in this normative analysis and demonstrates that the circumstances in which metadata is shared today – be it telephony, internet, location or even biometric “metadata” – are radically different from the circumstances of the cases that courts have relied upon to distinguish metadata from data thus far.¹ These differences primarily manifest themselves in the ability of information subjects to share information *voluntarily*; the ability of the holders of our metadata to aggregate, store, combine and analyze that data; and the extent to which we, the data subjects, *assume the risk* of our metadata being shared beyond the purpose for which we originally provided it. Significantly, we propose a three-pronged test for evaluating the voluntariness of sharing information with third parties, namely, first, whether a person *knowingly* shares information with a third party, second, whether a person has an *alternative* not to do so, and third, whether that alternative is *reasonable*. Adopting the framework of contextual integrity, the paper then assesses the impacts of social and technological changes in the information environment on the *actors, attributes* and *transmission principles* of relevant information flows to determine whether or not the NSA’s bulk telephony metadata collection program violates the principle of contextual integrity, and hence privacy expectations. Finally, our evaluation of the program in terms of contextual values and ends demonstrates that the benefits of the bulk collection of telephony metadata to national security are heavily outweighed by the program’s costs, including both the money and manpower invested in the program, on the one hand, and its infringements on civil liberties, such as privacy, freedom of speech and association, transparency, due process and the balance of power between the government and its citizens, on the other hand. The paper concludes that the prima facie assumption underlying the NSA’s justification for the bulk telephony metadata collection program – namely, that metadata is equivalent to non-sensitive data – no longer makes sense. Indeed, in light of the theory of contextual integrity, it never made any sense to begin with.

¹. Most notably Smith v. Maryland, 442 U.S. 735 (1979); see *infra* Section II.B.
II. ONTOLOGICAL ANALYSIS

A. Defining Metadata

Metadata is “data about data,” or information used to classify other information. Metadata has played an important role in library and computer sciences because it allows for knowledge management. Metadata such as the author, title, date of a publication, as well as publisher, size, amount of pages, and genre, for instance, enables a useful classification of books. In a library catalog, metadata may allow patrons to locate books they were expressly seeking or to discover books of potential interest about which they may not have previously known.

Metadata thus plays an important role in a myriad of different information searches. But metadata serves critical informational functions beyond that, often in contexts we do not usually associate with the term. A Word document, for example, includes such metadata as the file name and author, the time and date of its creation, and the file format and size. Tweets not only consist of 140 signs but also of a host of metadata, including the author’s name and biography, the date his account was created, the number of users he is following, and the location and time zone from which his Tweet is sent. What makes metadata so useful is the fact that, generally unlike the contents of a communication, it can be easily read and processed by machines. Indeed, as Cory Doctorow noted in 2001, if everyone were to “create good metadata for the purposes of describing their goods, services and information, it would be a trivial matter to search the Internet for highly qualified, context-sensitive results: a fan could find all the downloadable music in a

2. See Metadata Definition, DICTIONARY.COM, http://dictionary.reference.com/browse/metadata, (last visited June 23, 2015), defining metadata as “information that is held as a description of stored data”) and Metadata Definition, OED.COM, http://www.oed.com/view/Entry/117150?redirectedFrom=metadata (last visited June 23, 2015), defining metadata as “data that describes and gives information about other data”. The OED also quotes Philip R. Bagley’s definition of metadata: “As important as being able to combine data elements to make composite data elements is the ability to associate explicitly with a data element a second data element which represents data ‘about’ the first data element. This second data element we might term a ‘metadata element’. Examples of such metadata elements are: an identifier, a main ‘prescriptor’ which specifies from what domain the value of the first element must be taken, an access code which limits the conditions under which the first data element can be accessed,” PHILIP R. BAGLEY, EXTENSION OF PROGRAMMING LANGUAGE CONCEPTS 26 (1968).


given genre, a manufacturer could efficiently discover supplies, travelers could easily choose a hotel room for an upcoming trip.”

However, the term metadata gained some notoriety in the aftermath of the Snowden revelations as the first set of documents released demonstrated that a classified Foreign Intelligence Surveillance Court (FISC) order had compelled the American telecommunications company Verizon to hand over the telephony metadata of virtually all American subscribers on an ongoing permanent basis. Telephony metadata in this case included “communications routing information, including but not limited to session identifying information (e.g., originating and terminating phone number, International Mobile Subscriber Identity (IMSI) number, International Mobile station Equipment Identity (IMEI) number, etc.), trunk identifier, telephone calling card numbers, and time and duration of call” but not “the substantive content of any communication.”

In its defense, the NSA asserted that, because it was collecting only the metadata and not the content of communications, its bulk telephony metadata collection program did not raise any privacy concerns.

This argument has now been scrutinized in court. But while Judge Leon in Klayman v. Obama (2013) reached the conclusion that the program violated a reasonable expectation of privacy under the Fourth Amendment, Judge Pauley in ACLU v. Clapper (2013) ruled that it did not. After tracing the genealogy of cases upon which these respective decisions were based, we discovered that the courts, in attempting to classify the term metadata in light of precedent, relied on three constitutionally relevant distinctions, namely non-content v. content data; business records held by third parties v. private records; and information “in plain view” v. hard-to-obtain information. Although neither of the judges argued that the conception of metadata – its

ontological status – had changed over time, for Judge Pauley an *ontological* analysis of metadata was sufficient for resolving the case, whereas Judge Leon also engaged in a *normative* analysis of the evolving social and technological environment – an analysis which was ultimately critical to his decision.

**B. Classifying Metadata in the Law**

**Content v. non-content**

The first time the U.S. Supreme Court distinguished between content and non-content was in the context of *Ex parte Jackson* (1848), which questioned whether and the extent to which U.S. authorities could interfere with the mail in order to prevent the circulation of “obscene” materials. The Court concluded that “a distinction is to be made between different kinds of mail matter – between what is intended to be kept free from inspection, such as letters, and sealed packages subject to letter postage, and what is open to inspection, such as newspapers, magazines, pamphlets, and other printed matter purposely left in a condition to be examined. Letters and sealed packages of this kind in the mail are as fully guarded from examination and inspection, *except as to their outward form and weight*, as if they were retained by the parties forwarding them in their own domiciles.”

The Court thus argued that, while police officers should not be able to tear open sealed letters and packages, they should be free to look at the outside appearance and labeling thereof since that information necessarily had to be visible in order to enable them to deliver mail from origin to destination.

In 1928, by contrast, the Court ruled in *Olmstead* that the content of telephone conversations was not analogous to the content of sealed letters. It argued that

> It is plainly within the words of the Amendment to say that the unlawful rifling by a government agent of a sealed letter is a search and seizure of the sender’s papers or effects. The letter is a paper, an effect, and in the custody of a Government that forbids carriage except under its protection. The United States takes no such care of telegraph or telephone messages as of mailed sealed letters. The Amendment does not forbid what was done here. There was no searching. There was no seizure. The evidence was secured by the use of the sense of hearing and that only. There was no entry of the houses or offices of the defendants.

---

12. The protection of the content of sealed letters and packages does not extend to fourth class mail, however. *See United States v. Riley*, 554 F.2d 1282, 3-4 (1977) (arguing that “unlike first class mail, there is no expectation of privacy in the forwarding of fourth class mail” since the petitioner “could have availed himself of the protection afforded by first class postage”).
Justice Brandeis, however, disagreed: “the mail is a public service furnished by the Government. The telephone is a public service furnished by its authority. There is, in essence, no difference between the sealed letter and the private telephone message.” The content of telephone conversations should consequently enjoy the same legal protection as the content of letters.

The Court revisited this question four decades later in *Katz v. United States* (1967). In *Katz* the police had attached an electronic listening and recording device to the outside of a telephone booth from which the petitioner was conducting illegal gambling activities. The Court ruled that, although it had “supposed in *Olmstead* that surveillance without any trespass and without the seizure of any material object fell outside the ambit of the Constitution, we have since departed from the narrow view on which that decision rested. Indeed, we have expressly held that the Fourth Amendment governs not only the seizure of tangible items, but extends as well to the recording of oral statements.” According to the *Katz* Court, “no less than an individual in a business office, in a friend’s apartment, or in a taxicab, a person in a telephone booth may rely upon the protection of the Fourth Amendment. One who occupies it, shuts the door behind him, and pays the toll that permits him to place a call is surely entitled to assume that the words he utters into the mouthpiece will not be broadcast to the world. To read the Constitution more narrowly is to ignore the vital role that the public telephone has come to play in private communication.” But the *Katz* decision was exclusively about the protection of the contents of telephone conversations; the *Katz* Court was not in a position to comment on what protections extended to information that was solely about the call. Rather, the extent to which non-content as opposed to the content of telephone conversations would be protected by the Fourth Amendment was decided in the context of *United States v. New York Telephone Company* (1977) and *Smith v. Maryland* (1979).

In *New York Telephone Company*, the Court addressed the question of whether installing and using a pen register device to record the numbers dialed from a phone, falls under the definition of an “intercept” according to Title III of the Omnibus Crime Control and Safe Streets Act of 1968 (the Wiretap Statute). The Act defined an “intercept” as “the aural intercept of the contents of any wire or oral communication through the use of any electronic, mechanical, or

14. *Id.* at 475.
16. *Id.* at 352.
other device.”

The Court denied that pen registers fall under the definition of the Act “because they do not acquire the ‘contents’ of communications, as the term is defined by 18 U.S.C. 2510(8).”

As the Court explained, “these devices do not hear sound. They disclose only the telephone numbers that have been dialed – a means of establishing communication. Neither the purport of any communication between the caller and the recipient of the call, their identities, nor whether the call was even completed is disclosed by pen registers.”

In distinguishing the register of numbers dialed from the aural, substantive content of the call, the Court thus made an ontological distinction between data and metadata on which it furthermore based the normative conclusion that “Congress did not view pen registers as posing a threat to privacy of the same dimension as the interception of oral communications.”

The distinction between content and non-content information of telephone conversations was further solidified in Smith v. Maryland, a landmark case in Fourth Amendment doctrine. The Smith Court addressed the question of whether the warrantless installation and use of a pen register violated a reasonable expectation of privacy under the Fourth Amendment. The petitioner claimed that it did, drawing an analogy to the intercept at issue in Katz. However, the Smith Court distinguished pen registers “from the listening device employed in Katz, for pen registers do not acquire the contents of communications” and consequently ruled that pen register information fell outside the ambit of the Fourth Amendment. In response to Smith v. Maryland, Congress passed the Pen Register Act, which ensured that pen register information would be granted at least some form of protection under the law, albeit substantially weaker than that granted to the content of communications.

Because the distinction between content and non-content information is primarily drawn in the statutory and not the constitutional context, neither Judge Leon nor Judge Pauley explicitly discuss it in the context of Fourth Amendment law. Both judges do accept, however, that the metadata at issue in the NSA’s program is essentially equivalent to the pen register data at issue.

---

19. Id. at 167.
20. Id. at 168.
22. See DANIEL J. SOLOVE AND PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 295 (5th ed. 2015). The Pen Register Act does not, however, apply to the NSA’s bulk telephony metadata collection program because the latter is considered foreign intelligence collection and thus governed by the Foreign Intelligence Surveillance Act (FISA), as amended by Section 215 of the USA PATRIOT Act. See infra Section III.B. For a detailed statutory and constitutional analysis of the NSA’s bulk telephony metadata collection program, see Laura K. Donohue, Bulk Metadata Collection: Statutory and Constitutional Considerations, 37 HARVARD J.L. & PUB. POL’Y 757 (2014).
in *Smith*. As Judge Leon points out, “what metadata *is* has not changed over time. As in *Smith*, the *types* of information at issue in this case are relatively limited: phone numbers dialed, date, time, and the like.”23 Since the case law suggests that pen register data is non-content information, and non-content information is equivalent to non-sensitive information, the telephony metadata collected by the NSA was thus prima facie deemed to be non-sensitive information as well.

**Private records v. Business records held by third parties**

Another distinction that significantly influenced the decisions in *ACLU v. Clapper* and *Klayman v. Obama* is that between *private records* and *business records held by third parties*. This distinction goes back to a series of cases in the 1950s and 60s in which the Supreme Court decided that the protection of the Fourth Amendment did not extend to incriminating statements made in the presence of undercover police agents.24 According to the Court, a defendant does not have “a justifiable and constitutionally protected expectation that a person with whom he is conversing will not then or later reveal the conversation to the police.”25 Significantly, the scope of the doctrine of misplaced trust, initially limited to people, was eventually expanded to include businesses too. In *United States v. Miller* (1976), the Supreme Court ruled that the petitioner did not have a reasonable expectation of privacy in his bank records because they only contained “information voluntarily conveyed to the banks and exposed to their employees in the ordinary course of business.”26 Referencing *White*, the Court specified 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.”27 The majority in *Miller* thus established the consequential third-party doctrine. In *Smith*, too, the

---

23. 2013 Klayman Opinion, *supra* note 8 at 52. Quoted in 2013 Clapper Opinion, *supra* note 9 at 44. At the same time, Judge Leon acknowledges that his statement is not entirely accurate since “the pen register in *Smith* did not tell the government whether calls were completed or the duration of any calls, whereas that information is captured in the NSA’s metadata collection.” Furthermore, “telephony metadata can reveal the user’s location, which in 1979 would have been entirely unnecessary given that landline phones are tethered to buildings.” 2013 Klayman Opinion, *supra* note 8 at 52 n.57 (internal citations omitted).


27. *Id.* at 443.
Court accepted this reasoning and argued that “a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties.”

The third-party doctrine significantly influenced Judge Pauley’s decision in ACLU v. Clapper: “Clear precedent applies because Smith held that a subscriber has no legitimate expectation of privacy in telephony metadata created by third parties.” Although Judge Leon also acknowledges that the type of data collected by the government in its bulk telephony metadata program is essentially equivalent to the pen register data at issue in Smith, for him this fact was not dispositive. Rather, as we demonstrate in the following section, Judge Leon’s decision was influenced by a series of cases that were far more concerned with the changing nature and circumstances of government information collection than with the type of information collected.

**Hard-to-obtain information v. Information in plain view**

The distinction between hard-to-obtain information and information “in plain view” differs from the previous two cases in that it does not depend on the type of information at hand, but rather on how easily available it is. Generally, the courts have decided that petitioners have no reasonable expectation of privacy in information “in plain view” because they assumed the risk that the police would have access to that information. The “plain view doctrine” is based on

---

29. 2013 Clapper Opinion, supra note 9 at 44 (internal citations omitted).
30. See 2013 Klayman Opinion, supra note 8 at *43 (“The Supreme Court held that Smith had no reasonable expectation of privacy in the numbers dialed from his phone because he voluntarily transmitted them to his phone company.”)
31. See United States v. Lee, 274 U.S. 559 (1927) (arguing that the use of search lights to examine objects in plain view is permissible under the Fourth Amendment), Ker v. California 374 U.S. 23, 43 (1963) (arguing that the seizure of a brick of marijuana “did not constitute a search, since the officer merely saw what was placed before him in full view”), Lewis v. United States 385 U.S. 206, 211 (1966) (arguing that “when, as here, the home is converted into a commercial center to which outsiders are invited for purposes of transacting unlawful business, that business is entitled to no greater sanctity than if it were carried on in a store, a garage, a car, or on the street”), Katz v. United States, 389 U.S. 347, 351 (1967) (arguing that “[w]hat a person knowingly exposes to the public, even in his home or office, is not a subject of Fourth Amendment protection), Harris v. United States. 390 U.S. 234, 236 (1968) (arguing that “it has long been settled that objects falling in the plain view of an officer who has the right to be in the position to have that view are subject to seizure and may be introduced in evidence”). The “plain view doctrine” was extended by the “open fields doctrine” in that individuals have no reasonable expectation of privacy in the fields that they own. See Hester v. United States, 265 U.S. 57 (1924) (arguing that the Fourth Amendment does not apply to objects discarded in the open fields even if the land belongs to the petitioner), Oliver v. United States, 466 U.S. 170, 171 (1984) (arguing that “[b]ecause open fields are accessible to the public and the police in ways that a home, office, or commercial structure would not be, and because fences or ‘No Trespassing signs do not effectively bar the public from viewing open fields, the asserted expectation of privacy in open fields is not one that society recognizes as reasonable”). The courts did, however, carve out an exception for a home’s so-called curtilage because “the area
the idea that officers should not have to turn away from evidence that is right in front of their eyes. However, whether or not something is “in plain view” has become more complicated over time given that police officers now increasingly resort to advanced, technologically-enabled surveillance techniques. For instance, in *Florida v. Riley* (1989) the police circled the respondent’s property with a helicopter in order to discover, by virtue of two missing roof panels, that the respondent was growing marijuana in the greenhouse adjacent to his home. The majority argued that, “the police, like the public, would have been free to inspect the backyard garden from the street if their view had been unobstructed. They were likewise free to inspect the yard from the vantage point of an aircraft flying in the navigable airspace as this plane was.”

However, the dissent objected, saying that the question “must be not whether the police were where they had a right to be, but whether public observation of Riley’s curtilage was so commonplace that Riley’s expectation of privacy in his backyard could not be considered reasonable.” A similar question was raised in *Dow Chemical Co. v. United States* (1986) where the Environmental Protection Agency (EPA) used a precision aerial mapping camera to take photographs of a chemical plant. The majority argued that “[t]he mere fact that human vision is enhanced somewhat, at least to the degree here, does not give rise to constitutional problems.”

The dissent, on the other hand, objected that this would undermine the court’s longstanding “standard that ensured that Fourth Amendment rights would retain their vitality as technology expanded the Government’s capacity to commit unsuspected intrusions into private areas and activities.” The point at which advanced surveillance techniques become so intrusive that one can no longer speak of information “in plain view” can be determined somewhat more easily when the technologies in question capture information emanating from within the home. For instance, when in *Kyllo v. United States* (2001) the police used a thermal imager to determine whether or not the suspect was growing marijuana in his home the court argued that the officers obtained information that is “not visible to the naked eye” and thus intruded upon the constitutionally protected space of the home. In sum, the cases we just discussed were concerned with whether or not police officers were gathering information “in plain view”. Evidence “in

---

33. *Id.* at 460 (Brennan, J., joined by Marshall and Stevens, JJ., dissenting).
35. *Id.* at 240 (1986).
plain view” is not subject to Fourth Amendment protection; however, what constitutes information in “plain view” has been complicated by the introduction of new technologies. When, as in Kyllo, new technologies are used to capture information emanating from within the home, the Court is generally more willing to extend Fourth Amendment protection.

The court was confronted with a similar set of questions in cases involving information in public. In United States v. Knotts (1983), the police was able to follow the movements of the defendant by attaching a beeper to a chloroform container he had purchased and placed in his car. The respondent challenged the monitoring on Fourth Amendment grounds; however, the Supreme Court ruled that “[t]he beeper surveillance amounted principally to following an automobile on public streets and highways. A person traveling in an automobile on public thoroughfares has no reasonable expectation of privacy in his movements.”37 A beeper, the Court reasoned, did not raise Fourth Amendment concerns in this case because it did not reveal any information “that would not have been visible to the naked eye.”38 The respondent objected that the result of the ruling would be that “twenty-four hour surveillance of any citizen of this country will be possible, without judicial knowledge or supervision.”39 But the Court suggested that these were mere speculations and “if such dragnet-type law enforcement practices as respondent envisions would eventually occur, there will be time enough then to determine whether different constitutional principles may be applicable.”40 However, in United States v. Karo (1984), the Court specified that, while the attachment of a beeper itself does not raise any Fourth Amendment concerns, once it allows the police to track movements within a house it certainly does.41 Karo was thus distinguished from Knotts on the basis that the beeper in Knotts was used to monitor movements in public, the beeper in Karo used to monitor movements in a private space.42 In other words, in Knotts the defendant assumed the risk of surveillance when travelling on public thoroughfares, whereas in Karo the defendant invoked his right against government surveillance within the constitutionally protected space of the home.

At the same time, the court did not categorically rule out that a person could have a reasonable expectation of privacy in information that is theoretically public. Indeed, in U.S. Dep’t
of Justice v. Reporters Comm. for Freedom of Press (1989) the Supreme Court recognized the defendant’s right to privacy in information contained in public rap sheets because “the compilation of otherwise hard-to-obtain information alters the privacy interest implicated by the disclosure of that information.” The court thus effectively created a right to practical obscurity in information contained in public rap sheets. While the Court has generally been reluctant to extend the concept of practical obscurity to other contexts, it resurfaced, at least implicitly, in United States v. Jones (2012) where the police engaged in the warrantless GPS monitoring of a suspect’s car for a period of 28 days. While the majority resolved the case on the basis of the trespass doctrine, arguing that it was the attachment of the GPS device to the car that violated the defendant’s Fourth Amendment rights, Justices Sotomayor and Alito, in separate concurrences, questioned whether the aggregation of otherwise public information over time would not raise separate constitutional concerns. Most importantly, Justice Sotomayor raised the question of whether it might not 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. People disclose the phone numbers they dial or text to their cellular providers; the URLs that they visit and the e-mail addresses with which they correspond to their Internet service providers; and the books, groceries, and medications they purchase to online retailers.” In so doing, Justice Sotomayor effectively opened the door to future challenges to the third party doctrine established under Miller and Smith. Particularly noteworthy in the context of this paper, however, is that Justice Sotomayor also drew an analogy between the privacy concerns raised by the aggregation of information in

44. “Where, as here, the subject of a rap-sheet is a private citizen and the information is in the Government’s control as a compilation, rather than as a record of what the Government is up to, the privacy interest in maintaining the rap-sheet’s “practical obscurity” is always at its apex, while the FOIA-based public interest in disclosure is at its nadir.” DOJ, 489 U.S. at 750 (emphasis added). See also Danny Weitzner, Privacy, Practical Obscurity and the Power of the Semantic Web, MIT DECENTRALIZED INFORMATION GROUP (April 9, 2006, 11:44 PM), http://dig.csail.mit.edu/breadcrumbs/node/125 (defining practical obscurity as “legal doctrine that one may have a privacy interest in the compilation of information (aka a dossier) even though each piece of information composing the dossier is itself publicly available”).
plain view (such as a car on public thoroughfares) and information “voluntarily” provided to third parties (referring extensively to information generally classified as metadata).

Judge Pauley, however, rejected “the ACLU’s reliance on the concurring opinions in Jones” because ultimately the “Supreme Court did not overrule Smith.” For Judge Leon, on the other hand, the fact that the NSA was collecting telephony metadata in bulk and on an “ongoing daily basis” was dispositive to distinguishing Klayman v. Obama from Smith v. Maryland. For Judge Leon, the essential question was ultimately not ontological but normative, which is also why he decided, despite the similarity of the data at hand in Klayman and Smith, that

The question before me is not the same question that the Supreme Court confronted in Smith. To say the least, ‘whether the installation and use of a pen register constitutes a ‘search’ within the meaning of the Fourth Amendment’ – under the circumstances addressed and contemplated in that case – is a far cry from the issue in this case. Indeed, the question in this case can more properly be styled as follows: When do present-day circumstances – the evolutions in the Government’s surveillance capabilities, citizens’ phone habits, and the relationship between the NSA and telecom companies – become so thoroughly unlike those considered by the Supreme Court thirty-four years ago that a precedent like Smith simply does not apply? The answer, unfortunately for the Government, is now.

47. 2013 Clapper Opinion, supra note 9 at 43. In support of this argument, see Orin Kerr, Debate: Metadata and the Fourth Amendment – A Reply to Jennifer Granick, JUST SECURITY (September 23, 2013, 9:30 AM), https://www.justsecurity.org/1009/debate-metadata-fourth-amendment-reply-jennifer-granick/. 48. Greenwald, supra note 5. 49. 2013 Klayman Opinion, supra note 8 at 47. 50. 2013 Klayman Opinion, supra note 8 at 45. Laura K. Donohue advances a similar argument in comparing bulk telephony metadata and the pen register data at issue in Smith, see supra note 22 at 870-871 (2014) (“The extent to which we rely on electronic communications to conduct our daily lives is of a fundamentally different scale and complexity than the situation that existed at the time the Court heard arguments in Smith. Resultantly, the extent of information that can be learned about not just individuals, but about neighborhoods, school boards, political parties, Girl Scout troops – indeed, about any social, political, or economic network – simply by placement of a pen register or trap and trace, is far beyond what the Court contemplated in 1979.”) See also Jennifer Granick, Debate: Metadata and the Fourth Amendment, JUST SECURITY (September 23, 2013, 2:00 PM), https://www.justsecurity.org/927/metadata-fourth-amendment/. For a general overview of how the evolving social and technological environment affects, or should affect, Fourth Amendment law, see Katherine J. Strandburg, Home, Home on the Web and Other Fourth Amendment Implications of Technosocial Change, 70 MD. L. REV. 614 (2011).
In sum, while Judge Pauley emphasizes that the nature of telephony metadata has not changed (telephony metadata are business records, following *Smith*), Judge Leon in *Klayman v. Obama* argues that even if the nature of telephony metadata has not changed (and this is debatable), what has changed is the social and technological environment in which metadata is collected (the aggregation of bulk telephony metadata is not comparable to the limited collection of pen register information at issue in *Smith*, echoing *Jones*). In the normative analysis of our paper, we support and further expand on this claim. Based on the theory of contextual integrity, we provide a rigorous account why, even if one accepts that a reasonable line can be drawn between data and metadata, the prima facie assumption that data deserves greater privacy protections than metadata is fundamentally unsound.
III. NORMATIVE ANALYSIS

A. Contextual Integrity

According to the theory of contextual integrity, the appropriateness of a particular information flow depends not only on the type of information in question (the attribute) but also on the actors involved (senders, subjects and recipients of an information type) and the transmission principles (constraints on flow). If a practice generates changes in any of these three parameters, a prima facie case exists for claiming that contextual integrity, and hence privacy, has been violated. However, this prima facie assessment does not necessarily mean that the new practice needs to be abandoned. Indeed, if the new practice better promotes the values, goals and ends of a given context, then contextual integrity allows for and even encourages alterations in information flows.

Material advances in the science and technology of data as well as institutional practices have dramatically altered the social and technological environment in which metadata is generated and collected today. The framework of contextual integrity allows us to assess the impact that these changes have had on information flows. Accordingly, when information subjects no longer share their metadata voluntarily this affects the transmission principle; when the recipients of metadata have vastly increased capabilities of aggregating, storing, combining and analyzing metadata this changes the attribute; and when we assume the risk of surveillance whenever we impart with information, this introduces a wider range of actors into the information flow. A careful analysis of these interdependencies also challenges the dichotomies courts have used to distinguish metadata from data, namely content vs. non-content data, private records vs. business records held by third parties, and hard-to-obtain information vs. information “in plain view.” As we argue below, the fact that we no longer share information voluntarily undermines the notion that business records held by third parties deserve fewer privacy protections than private information held by the data subjects themselves; the fact that metadata is now aggregated, stored, combined and analyzed to enable a host of inferences to be drawn undermines the notion that metadata is non-content and therefore non-sensitive data; and the fact

52. Id. at 182 (2010). For a more detailed summary of the theory of contextual integrity, see Helen Nissenbaum, Respect for Context as a Benchmark for Privacy Online: What it is and isn’t, in SOCIAL DIMENSIONS OF PRIVACY 285-288 (Beate Roessler and Dorota Mokrosinska eds., 2015).
53. See infra Section II.B.
that most of our metadata is no longer hard to obtain undermines the notion that we should assume the risk of surveillance whenever we impart with it.\textsuperscript{54} We elaborate these claims below and explain their significance for our overarching thesis.

Table 1: Impacts of changes in the social and technical environment on information flows and respective normative implications

<table>
<thead>
<tr>
<th>Changes in the social and technical environment</th>
<th>Impact on information flow</th>
<th>Normative implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voluntariness</td>
<td>Voluntarily sharing information means knowingly sharing information and having a reasonable alternative not to do so. Even if telephone subscribers knowingly share their metadata with phone companies, they do not have a reasonable alternative not to do so. Telephone subscribers therefore do not share their metadata voluntarily.</td>
<td>This changes the transmission principle of the information flow</td>
</tr>
<tr>
<td>Capabilities</td>
<td>The recipients of our metadata have vastly increased capabilities of aggregating, storing, combining and analyzing that data.</td>
<td>This changes the attribute of the information flow</td>
</tr>
<tr>
<td>Assumption of risk</td>
<td>Previously hard to obtain metadata has become easily accessible so that we effectively assume the risk of surveillance whenever we impart with the information.</td>
<td>This introduces a vastly increased range of actors into previously limited information flows</td>
</tr>
</tbody>
</table>

**Voluntariness**

The question of voluntariness was already raised in the early days of the third-party doctrine. In United States v. Miller, the court rejected the idea that the respondent enjoyed Fourth Amendment protection in his bank records because “all of the documents obtained, including financial statements and deposit slips, contain only information voluntarily conveyed to the banks

\textsuperscript{54} The last point, in particular, may seem counterintuitive at first, but makes sense when placed in the larger context of a free and open society, in which we should not be forced to assume the risk of surveillance whenever we impart with information; especially when we do not have a reasonable choice to withhold it.
and exposed to their employees in the ordinary course of business.” Along the same lines, in Smith v. Maryland the majority suggested that “When he used his phone, petitioner voluntarily conveyed numerical information to the telephone company and ‘exposed’ that information to its equipment in the ordinary course of business, In so doing, petitioner assumed the risk that the company would reveal to police the numbers he dialed.” However, the notion that the petitioners “voluntarily” handed over their data to the phone company and bank, respectively, was already contested at the time the cases were argued. In Miller, for example, Justice Brennan noted that “the disclosure by individuals or business firms of their financial affairs to a bank is not entirely volitional, since it is impossible to participate in the economic life of contemporary society without maintaining a bank account.”

Compare these two with cases involving cell-site location information (CSLI):
Theoretically, CSLI is non-content information “in plain view” voluntarily conveyed to third parties, and some courts have indeed supported that claim. For instance, according to the United States Court of Appeals for the Fifth Circuit, “Because a cell phone user makes a choice to get a phone, to select a particular service provider, and to make a call, and because he knows that the call conveys cell site information the provider retains this information, and the provider will turn it over to the police if they have a court order, he voluntarily conveys his cell site data each time he makes a call.” But this distorts the original information flow. A cell phone user (the sender and information subject) conveys his cell site data (the attribute) “voluntarily” (the transmission principle) only to the service provider (the intended recipient). Once the information is passed to the police, the cell phone user has not “voluntarily” provided anything at all. Neither the recipient nor the transmission principle of the information flow is the same anymore.

A further complication in the CSLI cases arises from the fact that most cell phone users do not even provide CSLI knowingly. As noted in the ruling, “When a cell phone user makes a call, the only information that is voluntarily and knowingly conveyed to the phone company is the number that is dialed and there is no indication to the user that making that call will also

55. Miller, 425 U.S. at 442 (emphasis added).
56. Smith, 442 U.S. at 744 (emphasis added).
57. For an early criticism in the academic literature, see John S. Applegate and Amy Applegate, Pen Registers after Smith v. Maryland, 15 HARVARD CIV. RIGHTS – CIV. LIBERTIES LAW REV. 753, 765 (1980) (claiming that “The argument advanced by the Court that telephone users know that records will be made of toll calls and thus have no expectation of privacy is unconvincing.”)
59. In re: Application of the United States of America for Historical Cell Site Data, 724 F.3d 600, 614 (5th Cir. 2013).
locate the caller; when a cell phone user receives a call, he hasn’t voluntarily exposed anything at all. And a caller most certainly does not voluntarily provide the registration information that the phone automatically sends to the phone company every seven seconds whenever the phone is on, without notice to or control by the user.”

This sentiment was later echoed by the Eleventh Circuit and most recently by the Fourth Circuit. The U.S. District Court, Southern District of Texas, further explains:

Unlike the bank records in Miller or the phone numbers dialed in Smith, cell site data is neither tangible nor visible to a cell phone user. When a user turns on the phone and makes a call, she is not required to enter her own zip code, area code, or other location identifier. None of the digits pressed reveal her own location. Cell site data is generated automatically by the network, conveyed to the provider not by human hands, but by invisible radio signal. Thus, unlike in Miller and Smith, where the information at issue was unquestionably conveyed by the defendant to a third party, a cell phone user may well have no reason to suspect that her location was exposed to anyone. The assumption of risk theory espoused by Miller and Smith necessarily entails a knowing or voluntary act of disclosure; the Government has cited no case (and the court has found none) where unknowing, inadvertent disclosure of information by a defendant thereby precluded Fourth Amendment protection of that information.”

It is worth noting the connection between these insights and traditional philosophical positions on (moral) responsibility, which require, at the very minimum, that actors are morally responsible for their actions insofar as they have been performed freely and knowingly, that is, with comprehension. This account forms the basis of notions of informed consent and of course the foundation for legal concepts surrounding liability. For our purpose, the general level of understanding of digitally intermediated communication for the most part may be insufficient as the basis for claiming voluntary, let alone knowing disclosure.

But even if the defendants in Miller and Smith voluntarily and knowingly conveyed information to a third party, they conveyed that information to a particular business for a particular purpose, and not to the police for the purpose of a criminal investigation. This is also one of the reasons why transforming personal records into business records should not diminish

---


63. See e.g. Joel Feinberg, Sua Culpa, in DEBORAH G. JOHNSON AND JOHN W. SNAPPER, ETHICAL ISSUES IN THE USE OF COMPUTERS (1st ed. 1985); JOHN MARTIN FISCHER AND MARK RAVIZZA, SJ., RESPONSIBILITY AND CONTROL: A THEORY OF MORAL RESPONSIBILITY (1st ed. 1998).
but rather reinforce the privacy expectations enjoyed therein. When personal information is provided to create a business record, the reasonable expectation is precisely that that information will only be used for a business purpose. The notion that there is no reasonable expectation of privacy in information held in business records is thus fundamentally flawed.\(^6^4\)

Several states have already rejected the third-party doctrine on this basis.\(^6^5\) For instance, in the context of a CSLI case in New Jersey, the state’s Supreme Court pointed out that “an individual’s privacy interests under New Jersey law does not turn on whether he or she is required to disclose information to third-party providers to obtain service. Just as customers must disclose details about their personal finances to the bank that manages their checking accounts, cell-phone users have no choice but to reveal certain information to their cellular provider. That is not a voluntary disclosure in a typical sense; it can only be avoided at the price of not using a cell phone.”\(^6^6\)

Similarly, the Supreme Court ruled in \textit{Ferguson v. City of Charleston} (2001) that the police could not conduct warrantless and nonconsensual drug tests on urine samples provided by pregnant women to a state hospital for the purpose of obstetric tests. According to the majority, the urine tests were “indisputably searches within the meaning of the Fourth Amendment” given that “none of the women searched provided either probable cause to believe that they were using cocaine, or even the basis for a reasonable suspicion of such use.”\(^6^7\) Moreover, under the two-pronged \textit{Katz} test, the majority ruled that “the reasonable expectation of privacy enjoyed by the typical patient undergoing diagnostic tests in a hospital is that the results of those tests will not be shared with nonmedical personnel without her consent.”\(^6^8\)

---

\(^6^4\). See also Kiel Brennan-Marquez, \textit{Fourth Amendment Fiduciaries}, 84 FORDHAM L. REV. 611, 654 (2015) (arguing that entrusting so-called “information fiduciaries” with personal information “carries an implicit limitation on use: specifically, an implied covenant to avoid using sensitive information in ways that harm the sharing party”).


\(^6^6\). State of New Jersey v. Thomas W. Earls, 70 A.3d 630, 641 (N.J. 2013). A similar argument has been made with regard to telephony and Internet metadata collection: “Even if U.S. citizens wanted to opt out of having this information collected, it would be virtually impossible to do so. There have, for instance, been advances in encryption. But these technologies all revolve around content – not metadata. Although some technologies are focused on metadata, these are not sufficiently advanced to allow for real-time communication. The only option is therefore not to use a telephone. The cost of doing so, however, would lean towards divesting oneself of a role in the modern world – impacting one’s social relationships, employment, and ability to conduct financial and personal affairs.” See Donohue, \textit{supra} note 22 at 874.


\(^6^8\). \textit{Id.} at 78 (emphasis added).
Justice Scalia acknowledges that “abuse of trust is surely a sneaky and ungentlemanly thing, and perhaps there should be (as there are) laws against such conduct by the government.” At the same time Scalia found that, until Ferguson v. City of Charleston, the majority has “never held – or even suggested – that material which a person voluntarily entrusts to someone else cannot be given by that person to the police, and used for whatever evidence it may contain. Without so much as discussing the point, the Court today opens a hole in our Fourth Amendment jurisprudence, the size and shape of which is entirely indeterminate.” Justice Scalia is correct in that the decision in Ferguson v. City of Charleston “represents a significant departure from the third-party doctrine. Indeed, is urine voluntarily turned over to a hospital really any different from tax documents turned over to a bank or metadata transmitted to the phone company?” However, most importantly Ferguson v. City of Charleston further undermines the third-party doctrine’s notion of “voluntariness” – because if the warrantless search of urine tests provided to a hospital for the purpose of obstetric test were constitutional it could only be avoided at the price of not making use of obstetric tests at all.

A Three-Pronged Test

In line with these conclusions, we suggest that courts should transform their assessment of “voluntariness” into a three-pronged test: first, whether a person knowingly shared information with a third party; second, whether a person had an alternative not to do so; and third, whether that alternative was reasonable. If the answer is “yes” to all questions (as is the case with misplaced trust in undercover police agents), then a person should not have a reasonable expectation of privacy in the information thus revealed. If the answer is “no” to all questions (as is the case with CSLI) or “yes” to the first two questions but “no” to the third (as is the case in Ferguson v. Charleston), then a person should retain a reasonable expectation of privacy in the information thus revealed.

69. Id. at 94 (Scalia, J., dissenting).
70. Id. at 95.
Capabilities

The significantly more limited collection of pen register data at issue in *Smith v. Maryland* in 1979 is not comparable to the *aggregation* of telephony metadata at issue in the bulk telephony metadata collection program, as conducted by the NSA in 2013. The police in *Smith* had ordered the telephone company to register the numbers dialed from the phone of a single person in the context of a specific, temporally limited police investigation. By contrast, the Snowden documents revealed that the NSA collected the telephony metadata of virtually all American subscribers in the context of an ongoing national security operation, regardless of whether or not they were suspected of any criminal behavior. Furthermore, as the *Klayman* court points out, the NSA program involved “the creation and maintenance of a historical database containing five years’ worth of data”\(^\text{72}\) and, at the time of the ruling, no end to the bulk telephony metadata collection program was envisioned either. According to the United States Court of Appeals for the Second Circuit, which reviewed Judge Pauley’s decision in *ACLU v. Clapper* and ultimately overruled it, “[s]uch expansive development of government repositories of formerly private records would be an unprecedented contraction of the privacy expectations of all Americans.”\(^\text{73}\)

Table 2: *A contextual comparison of Smith v. Maryland and the NSA’s bulk telephony metadata collection program*

<table>
<thead>
<tr>
<th></th>
<th>Smith v. Maryland (1979)</th>
<th>Bulk Telephony Metadata Collection Program (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td>Single police investigation</td>
<td>Ongoing national security investigation</td>
</tr>
<tr>
<td><strong>Senders</strong></td>
<td>Single American telephone subscriber suspected of criminal activity</td>
<td>All American telephone subscribers, regardless of whether or not they are suspected of any criminal activity</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>Various</td>
<td>Various</td>
</tr>
<tr>
<td><strong>Attribute</strong></td>
<td>Numbers dialed on a phone</td>
<td>Numbers that placed and received the call, the data, time, and duration of the call, other session-identifying information (for example, International Mobile Subscriber Identity number, International Mobile station Equipment Identity number, etc.), trunk identifier, and any telephone calling card number</td>
</tr>
<tr>
<td><strong>Transmission principle</strong></td>
<td>None; then ECPA (1986)</td>
<td>Section 215 of the USA PATRIOT Act (2001)</td>
</tr>
</tbody>
</table>

\(^{72}\) 2013 Klayman Opinion, *supra* note 8 at 47.

Furthermore, technological innovations also allow for a much greater storage of data. These innovations go both ways: not only can the NSA store a much larger amount of data, but individuals can also create and maintain much larger databases of personal information themselves. For instance, as the Supreme Court recognized in Riley v. California (2014), cell phones have effectively become “minicomputers”: “The current top-selling smart phone has a standard capacity of 16 gigabytes (and is available with up to 64 gigabytes). Sixteen gigabytes translates to millions of pages of text, thousands of pictures, or hundreds of videos.”

The amount of information that can be gleaned from a single device is unprecedented: “The sum of an individual’s private life can be reconstructed through a thousand photographs labeled with dates, locations, and descriptions [i.e., the metadata]; the same cannot be said of a photograph or two of loved ones tucked into a wallet.”

Cell phones also raise the specter of combination as they simultaneously function as “cameras, video players, rolodexes, calendars, tape recorders, libraries, diaries, albums, televisions, maps, or newspapers.” Cell phones nowadays contain not only call, but also Internet browsing history and location data, which, according to precedent, are equally considered non-content information held by third parties and information “in plain view,” respectively, and thus equally undeserving of the protection of the Fourth Amendment on a purely ontological basis.

One could object that the NSA did not actually collect all this information in its bulk telephony metadata collection program; however, as Judge Leon points out in several footnotes to his ruling, the exact scope of the telephony metadata program remains unclear. For instance, Judge Leon could not determine “whether ‘telephony metadata’ and ‘comprehensive communications routing information’ include data relating to text messages. If it does, then in 2012, the Government collected an additional six billion communications each day (69,635 each second).” Furthermore, despite the fact that later FISC orders explicitly prohibited the production of CSLI, “not all FISC orders have been made public, and I have no idea how location data has been handled in the past.” But even if the NSA did not collect these kinds of information as part of its bulk telephony metadata collection program – and the courts were

74. See also Felten Declaration, at 23 (emphasis added).
76. Id.
77. Id.
78. See United States v. Forrester, 512 F.3d 500 (9th Cir. 2008).
79. 2013 Klayman Opinion, supra note 8 at 54 n.56 (internal citations omitted).
80. Id.
therefore not in a position to define them – it is important to keep in mind that they could plausibly be defined as metadata in the future,\textsuperscript{81} which has serious privacy implications if judges continue to rely on an ontological rather than a contextual analysis of metadata collection.\textsuperscript{82}

In the context of Internet communications, that determination has already been made: In \textit{United States v. Forrester} (2007), the court ruled that Internet users have no reasonable expectation of privacy in e-mail headers and IP addresses because they should know that Internet service providers (ISPs) necessarily have access to that information in order to provide their services.\textsuperscript{83} Significantly, the court also found that the \textit{type} of information at hand was indistinguishable from the pen register information at issue in \textit{Smith}: “e-mail to/from addresses and IP addresses constitute addressing information and do not necessarily reveal any more about the underlying contents of communication than do phone numbers.”\textsuperscript{84} Based on a purely ontological analysis, addressing information is metadata and therefore non-sensitive data.

Finally, computers can easily \textit{analyze} metadata because it is structured and predictable. This distinguishes it from the content of communications which computers still struggle to comprehend.\textsuperscript{85} An analysis of the aggregated metadata then not only enables the NSA to infer patterns about social relations and associations,\textsuperscript{86} but also about political and religious beliefs.\textsuperscript{87}

\begin{flushleft}
\textsuperscript{81} The D.C. Court of Appeals already predicted in reviewing \textit{ACLU v. Clapper}: “If the government is correct, it could use §215 to collect and store in bulk any other existing metadata available anywhere in the private sector, including metadata associated with financial records, medical records, and electronic communications (including e-mail and social media information), relating to all Americans.” See 2015 Clapper Opinion at 818.

\textsuperscript{82} As a thought experiment, consider recent developments in the field of biometrics: FaceIt, a facial recognition software, “can pick someone’s face out of a crowd, extract the face from the rest of the scene and compare it to a database of stored images.” See Kevin Bonsor and Ryan Johnson, How Facial Recognition Systems Work, HOW STUFF WORKS (September 4, 2001), http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/facial-recognition1.htm. It can differentiate one face from another based on so-called nodal points, such as the distance between the eyes, width of the nose, depth of the eye sockets, the shape of the cheekbones, the length of the jaw line, etc. One could argue that the nodal points constitute the metadata of the face, whereas the identity of the person constitutes the content. Emotion sensing machines such as Affdex function in a similar manner: “The software scans for a face, if there are multiple faces, it isolates each one. It then identifies the face’s main regions – mouth, nose, eyes, eyebrows – and it ascribes points to each, rendering the features in simple geometries.” See Raffi Khatchadourian, We Know How You Feel, NEW YORKER, Jan. 15, 2015, at 51. The identifiers, again, could reasonably be described as the metadata of the face, whereas the emotions people are experiencing would be considered content. Since the latter can automatically be derived from the former, however, we imagine that people would demand an equal level of privacy protection for both.

\textsuperscript{83} Forrester, 512 F.3d at 510.

\textsuperscript{84} Id.

\textsuperscript{85} Especially spoken communication, given different rhythms and intonations of speech, as well as accents. See Felten Declaration, at 21.

\textsuperscript{86} For a visual representation of what the analysis of aggregated metadata might look like, see MIT’s Immersion project at https://immersion.media.mit.edu.

\textsuperscript{87} See Felten Declaration, at 46. Since the bulk collection of metadata enables law enforcement to identify networks and relationships, it also raises First Amendment, and in particular freedom of association, concerns. See Katherine J.
and even sensitive medical conditions. This is one of the big ironies behind the idea that metadata can be distinguished from data on the basis that it does not reveal anything about the underlying “content” of communications; indeed, as we outlined in Section II.A of our paper, the entire purpose of metadata in library and computer sciences is to classify and thus reveal essential aspects about the data that it describes.

What the foregoing analysis attempted to demonstrate is that, as capabilities in aggregation, storage, combination and analysis of metadata increase, so does the amount of information that can be gleaned from collecting it. We can reasonably expect that a cell phone service provider collects the numbers dialed from a phone. We generally do not expect that the provider simultaneously learns about our daily whereabouts, our friends and family, our professional associations, and our religious denomination. Therefore, as the social and technological environment changes, so might the attribute in an information flow. This is an important point because it fundamentally undermines one of the most important assumptions underlying the NSA’s justifications for the bulk telephony metadata program, namely that metadata is inherently non-sensitive data. From the perspective of the theory of contextual integrity, this assumption never made any sense to begin with; no information type is inherently sensitive or not. Rather, the privacy interest associated with a type of information can only be determined in light of an evaluation that takes into consideration all contextual parameters, including the senders, subjects and recipients as well as transmission principles governing the information flow.

---

\(^{88}\) Strandburg, Membership Lists, Metadata, and Freedom of Association’s Specificity Requirement, 10 ISILP 327 (2014).

\(^{88}\) Jonathan Mayer and Patrick Mutchler. Metaphone: The Sensitivity of Telephone Metadata, WEB POLICY (March 12, 2014), http://webpolicy.org/2014/03/12/metaphone-the-sensitivity-of-telephone-metadata/. Interestingly, a number of legal scholars already made this argument in the aftermath of Smith v. Maryland, see Applegate and Applegate, supra note 57, at 766. But of course the revelatory power of metadata today is even greater.

\(^{89}\) Sometimes metadata can be even more revealing than the content of communications: “Significant social analysis can also be conducted on the data. Sophisticated algorithms, for instance, can be applied to pen register information to ascertain where the important nodes are in a network. Alliances, friendships, and predilections can be uncovered by studying patterns in behavior. And unlike raw content, the type of information that can be gleaned is ordered – making it in some ways even more useful than the content itself.” See Donohue, supra note 22 at 871.
Assumption of risk

We often do not share our metadata voluntarily. Furthermore, technological innovations in aggregation, storage, combination and analysis increase the ability of those with whom we share our data to extract useful information from that data. But beyond that, changes in the social, technological and legal environment have made previously hard to obtain metadata easily accessible so that we effectively assume the risk of surveillance whenever we impart the information. The extension of the doctrine of misplaced trust from people to businesses is a telling case. Of course, nobody can protect us from sharing personal information with a friend who turns out to be not that great of a friend or, in a particular subset of cases, an undercover police agent.90 When we share intimate information with others we indeed assume the risk that those with whom we share our data will violate the terms of the social contract based on which this information was shared. However, the nature of the information sharing changes radically when the intended recipients of an information flow are no longer people but businesses: banks, telephone and web service providers, hospitals. The question here is no longer about the terms of a friendship but the terms of a transaction. Of course, the Supreme Court ruled in United States v. Miller (1977) 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.”91 However, in order to support this claim, the Court exclusively referenced cases that involve the use of undercover police agents, despite the fact that a relationship between people is governed by a different set of transmission principles than a relationship between a person and a bank.92 Justice Brennan readily dismissed this

90. In Germany, by contrast, undercover policing “entails a warrant procedure, a showing of need, and statutory limits on the crimes that the government may target in this way,” out of a historical concern for human dignity. See Jacqueline E. Ross, The Place of Covert Surveillance in Democratic Societies: A Comparative Study of the United States and Germany, 55 AM. J. COMP. L. 493, 562 (2007). Quoted in SOLOVE AND SCHWARTZ, supra note 22 at 287.

91. Miller, 425 U.S. at 443 (emphasis added). What makes the Miller case particularly contentious is that that “the bank did not just happen to be holding the records the government sought. Instead, the Bank Secrecy Act required (and continues to require) banks to maintain a copy of every customer check and deposit for six years or longer. The government thus compelled the bank to store the information, and then sought the information from the bank on the basis that since the bank held the data, there could not be any reasonable expectation of privacy and the Fourth Amendment therefore did not apply,” Fred H. Cate and Beth E. Cate, The Supreme Court and Information Privacy, 2 INTERNATIONAL DATA PRIVACY LAW 255, 263 (2012).

92. This discrepancy has been pointed out before: “Miller based this ‘assumption of risk’ argument on two informer cases, United States v. White and Hoffa v. United States. But to name these cases suggests the distinction: one expects a human being to evaluate, digest, recall, and perhaps repeat information; a bank merely performs and registers a transaction,” Applegate and Applegate, supra note 57, at 756. See also Brennan-Marquez, supra note 64.
argument in dissent, quoting representatives from several banks according to whom “a bank customer’s reasonable expectation is that, absent compulsion by legal process, the matters he reveals to the bank will be utilized by the bank only for internal banking purposes.” The dissent in Smith v. Maryland (1979) echoed similar concerns: “Those who disclose certain facts to a bank or phone company for a limited business purpose need not assume that this information will be released to other persons for other purposes.” Furthermore, “[a]t least in the third-party consensual surveillance cases, which first incorporated risk analysis into Fourth Amendment doctrine, the defendant presumably had exercised some discretion in deciding who should enjoy his confidential communications. By contrast here, unless a person is prepared to forgo use of what for many has become a personal or professional necessity, he cannot help but accept the risk of surveillance.”

Of course, sharing personal information and the associated metadata with third parties has become an even greater personal and professional necessity today. A 2013 longitudinal study of American youth revealed that young adults communicate about as much using digital media such as emails and social networks as they communicate face-to-face. Cell phones, as Judge Leon points out in Klayman v. Obama, have become ubiquitous: “Count the phones at the bus stop, in a restaurant, or around the table at a work meeting or any given occasion. Thirty-four years ago, none of these phones would have been there. Thirty-four years ago, city streets were lined with pay phones. Thirty-four years ago, when people wanted to send ‘text messages,’ they wrote letters and attached postage stamps.” The pervasive use of electronic devices, moreover often connected to the Internet, generates an overwhelming amount of metadata, all of which remains unprotected on the basis that we assume the risk of surveillance as soon as we disclose personal information to third parties. Most importantly, as Judge Leon points out in direct reference to

94. Smith, 442 U.S. at 749 (1979). The same argument was made in the context of mail covers: “[A] reasonable person expects (1) that the information contained in the return address will only be used for mail purposes, and (2) that it will be utilized in only a mechanical fashion without any records being kept. The recording and disclosure to non-postal authorities for non-postal purposes that results from a mail cover extends far beyond these narrow bounds,” United States v. Choate, 422 F. Supp. 261, 270 (C.D. Cal. 1976).
95. Smith, 442 U.S. at 749-750 (1979). A similar argument was made in the case preceding Smith v. Maryland: “Even if the majority’s analogy to Miller is valid, (and I do not agree) and Smith should have expected that the telephone company could itself monitor his phone for billing purposes, to improve service to its customers, or to verify complaints, Smith nevertheless had a reasonable expectation that the telephony company would not, without the safeguards of appropriate legal process, act for the government in collecting information relevant to a criminal prosecution,” Smith v. State, 389 A.2d 858 (Md. 1978) (Cole, J., dissenting).
97. 2013 Klayman Opinion, supra note 8 at 52.
U.S. Dep’t of Justice v. Reporters Comm. for Freedom of Press, “[i]t’s one thing to say that people expect phone companies to occasionally provide information to law enforcement; it is quite another to suggest that our citizens expect all phone companies to operate what is effectively a joint intelligence operation with the Government.” The NSA program thus essentially introduced a completely different set of recipients into the original information flow between telephone users and service providers, thereby violating contextual integrity and hence also the privacy expectations of the customers so affected.

Apart from U.S. Dep’t of Justice v. Reporters Comm. for Freedom of Press and Jones, the position that we should not be forced to assume the risk of surveillance whenever we impart with information, especially if that information was previously hard to obtain, was also supported in the context of financial privacy. A dispute arose between the Federal Trade Commission (FTC) and the credit unions when Congress passed the Gramm-Leach-Bliley Act (GLBA), which “required financial institutions to provide notice to consumers prior to transmitting covered information to others, and to permit them to opt out (subject to certain exceptions). This meant that credit header information, which previously had been freely sold (for such purposes as target marketing), was now subject to GLBA requirements.”

According to TransUnion and the Individual Reference Services Group (IRSG), only financial information could be regulated by the GLBA. However, in its definition of non-public information (NPI) Congress also included “any list, description, or other grouping of consumers (and publicly available information pertaining to them) that is derived using any nonpublic personal information other than publicly available information” 15 U.S.C. §6809(4)(C)(i). According to the court this meant that Congress had “provided for especially broad privacy protections for all information contained in these lists of consumers that is derived using nonpublic personal information. This is so even where the information is otherwise publicly available: the information is still protected, as long as it was derived using nonpublic personal information.”

The court thus recognized a right to privacy in the information, despite of the fact that it was available in public, because public information, when compiled and used out of context, violates a reasonable expectation of privacy.

98. 2013 Klayman Opinion, supra note 8 at 48.
99. NISSENBAUM, supra note 51, at 154.
100. Individual Reference Services Group, Inc. v. FTC, 145 F.Supp.2d 6, 28 (emphasis added).
B. Evaluation

The NSA’s bulk telephony metadata collection program has introduced changes in the actors, attributes and transmission principles of the information flows in a national security context and therefore constitutes a prima facie violation of the principle of contextual integrity. But the analysis does not end here. While the theory presumptively favors protecting the integrity of entrenched informational norms, it also allows for information environments to evolve “if new practices are demonstrably more effective at achieving contextual values, ends, and purposes or the equivalent.”\(^{101}\) A contextual analysis of the NSA’s bulk telephony metadata program therefore requires an evaluation of the moral and political factors affected by the bulk telephony metadata program and whether or not the benefits of altering information flows in this way justifies potential costs in light of contextually specific goals and ends.\(^{102}\)

The context in which the bulk telephony metadata collection program was conducted is that of national security. Following the traumatic events of September 11, 2001, the primary goal of US homeland security became the prevention of any further terrorist attacks and the concomitant loss of American lives on American soil. However, as several commentators pointed out, the intelligence community was unable to prevent the terrorist attacks of September 11, 2001 not because of insufficient information collection but because the FBI and NSA had an insufficient understanding of the rules that governed information sharing between intelligence agencies – information they already had thanks to conventional law enforcement techniques.\(^{103}\) Nevertheless, the experience of 9/11 left a lasting mark on members of the intelligence community in that they felt pressured to do everything in their power, and collect and share every piece of information they could get their hands on, in order to prevent a comparable terrorist attack from happening in the future.\(^{104}\)

The mindset of wanting to collect and share any kind of information even remotely connected to terrorism also goes a long way in explaining the intelligence community’s interpretation of the FISA business records provision on which it relied to justify the bulk telephony metadata collection program in statutory terms. The business records provision states

---

101. NISSENBAUM, supra note 51, at 180.
102. Id. at 182.
that “the Director of the Federal Bureau of Investigation or a designee of the Director (whose rank shall be no lower than Assistant Special Agent in Charge) may make an application for an order requiring the production of any tangible things (including books, records, papers, documents, and other items)” provided that “the tangible things sought are relevant to an authorized investigation.”\textsuperscript{105} Given that terrorists could be hiding behind any number of telecommunications devices in the United States, the government relied on this language to argue “that all telephone calls in the United States, including those of a wholly local nature, are ‘relevant’ to foreign intelligence investigations.”\textsuperscript{106} But as several commentators pointed out, this stretches the relevance standard beyond recognition as “any data might be ‘relevant’ to an investigation eventually, if by ‘eventually’ you mean ‘sometime before the end of time’.”\textsuperscript{107}

Theoretically these measures could, however, still be justified in light of contextual integrity if they ultimately better achieve the goal of national security, namely the prevention of future terrorist attacks. And indeed, the US government and intelligence representatives repeatedly claimed that the bulk telephony metadata collection program contributed to thwarting over 50 different terrorist attacks.\textsuperscript{108} However, this number has been scrutinized and ultimately debunked. The Privacy and Civil Liberties Oversight Board (PCLOB), which was tasked with evaluating the surveillance programs made public through the Snowden revelations, could not identify “a single instance involving a threat to the United States in which the program made a concrete difference in the outcome of a counterterrorism investigation.”\textsuperscript{109} The only case in which the bulk collection of telephony metadata played a significant role in the containment of terrorist activity was in the arrest of Basaaly Moalin, a Somali-born citizen, who was convicted of sending $8,500 to the Shabaab.\textsuperscript{110} However, even in that case, the PCLOB cautioned that “the

\textsuperscript{105} See 50 U.S.C. § 1861, emphasis added.
\textsuperscript{106} See Donohue, supra note 22 at 836-837 (2014).
\textsuperscript{110} Schwartz, supra note 104.
suspect was not involved in planning a terrorist attack and there is reason to believe that the FBI may have discovered him without the contribution of the NSA’s program.\footnote{111}

At the same time, the bulk collection of telephony metadata incurs significant costs. While a full assessment of the financial impact of the program is difficult because the information necessary to conduct such an evaluation remains classified, the program undoubtedly forms part of the “massive increases in homeland security expenditures that have taken place since 9/11 – increases that total well over $1 trillion.”\footnote{112} But the costs are not only material. Allocating resources to expand the surveillance apparatus in breadth necessarily forces the NSA to withdraw resources from conducting traditional and targeted surveillance techniques in depth – techniques that seem to be overall more effective at preventing terrorist attacks.\footnote{113}

Most importantly, as outlined above, the NSA’s bulk telephony metadata program fundamentally impacts civil liberties such as privacy, freedom of speech and association, transparency, due process and the balance of power between the government and its citizens. Ultimately, the bulk telephony metadata collection program not only has to be evaluated in the context of national security, but also in the larger context of the values of a free, open and above all democratic society. The evidence at hand suggests that the benefits provided by the program have not been significant enough to justify these costs.\footnote{114}

C. Outlook

The decisions in \textit{ACLU v. Clapper} and \textit{Klayman v. Obama} have both been reviewed by appeals courts. In \textit{ACLU v. Clapper}, the U.S. Court of Appeals for the Second Circuit vacated Judge Pauley’s decision, arguing that the bulk telephony metadata program went beyond the statutory authorization of §215.\footnote{115} In \textit{Klayman v. Obama}, on the other hand, the U.S. Court of Appeals for the D.C. Circuit overruled Judge Leon’s injunction on the basis that “the plaintiffs had not met the ‘higher burden of proof required for a preliminary injunction’ with regard to their

\footnotesize{111}. Privacy and Civil Liberties Oversight Board, \textit{supra} note 109.
\footnotesize{113}. See Bergen, Serman, Schneider, and Cahall, \textit{supra} note 103.
\footnotesize{114}. For an alternative cost-benefit analysis of the NSA’s bulk telephony metadata collection program, see Susan Freiwald, \textit{Nothing to Fear or Nowhere to Hide: Competing Visions of the NSA’s 215 Program}, 12 Colo. Tech. L.J. 309 (2014).
\footnotesize{115}. \textit{See} 2015 Clapper Opinion.
standing.” Both cases have been remanded to the respective district courts. Meanwhile Congress passed the USA Freedom Act, which, among other things, prohibits the bulk collection of telephony metadata by the NSA. In future, phone companies rather than the NSA will retain the metadata of their customers. The NSA can access it with court approval. However, neither the passage of the USA Freedom Act, nor the appeals decisions in ACLU v. Clapper and Klayman v. Obama systematically address the question of how Fourth Amendment challenges to bulk metadata collection programs will be handled in the future – an increasingly important question as the amount and diversity of metadata increases.

IV. CONCLUSION

This paper has demonstrated that a reasonable expectation of privacy depends not only on what metadata is – an ontological assessment – but also on the context in which it is created and collected – a normative assessment. The paper has shown that the circumstances in which metadata is shared today are radically different from the circumstances in which the distinctions between metadata and data were originally drawn. These differences primarily manifest themselves in the ability of information subjects to share information voluntarily; the ability of the recipients of our metadata to aggregate, store, combine and analyze that data; and the extent

118. Whether or not this improves the privacy protections of Americans remains to be seen. For instance, the European Court of Justice (ECJ) overruled the EU Data Retention Directive on the basis, among others, that it was unclear whether communications providers could meet “the obligations guaranteeing data protection and security” thus imposed on them. The ECJ further ruled that “the collection and, above all, the retention in huge databases, of the large quantities of data generated or processed in connection with most of the everyday electronic communications of citizens of the Union constitute a serious interference with the privacy of those individuals, even if they only establish the conditions allowing retrospective scrutiny of their personal and professional activities. The collection of such data establishes the conditions for surveillance which, although carried out only retrospectively when the data are used, none the less constitutes a permanent threat throughout the data retention period to the right of citizens of the Union to confidentiality in their private lives. The vague feeling of surveillance created raises very acutely the question of the data retention period.” Digital Rights Ireland Ltd and Kärntner Landesregierung and Others, Joined Cases C-293/12 & C-394/12, at 76, 72 (December 12, 2013) (Opinion of Advocate General Cruz Villalón).
to which we *assume the risk* of our metadata being shared beyond the purpose for which we originally provided it. Significantly, the paper proposes a three-pronged test for evaluating the voluntariness of third-party information sharing, namely, first, whether a person *knowingly* shares information with a third party, second, whether a person has an *alternative* not to do so, and third, whether that alternative is *reasonable*. On the basis of the theory of contextual integrity, the paper then analyzed how fundamental changes in the social and technological environment have affected the *actors, attributes* and *transmission principles* of relevant information flows, and concluded that the NSA’s bulk telephony metadata collection program violates the principle of contextual integrity, and hence privacy expectations. An evaluation of the program in light of contextually specific values and ends demonstrated that the costs incurred by the collection of telephony metadata in bulk – both in terms of the allocation of resources in the context of national security *and* in terms of the fundamental civil liberties affected by the program, such as privacy, freedom of speech and association, transparency, due process and the balance of power between the government and its citizens – ultimately do not justify its marginal benefits. Finally, the paper demonstrated that the prima facie assumption underlying the NSA’s program – namely, that metadata is equivalent to non-sensitive data – no longer makes sense. In light of the theory of contextual integrity, it never made any sense to begin with.