Consumer Privacy

Personally Identifiable Information

There is no uniform definition of ''Personally Identifiable Information,'' or PII, in the United States, the authors say. For privacy law to remain effective in the future, they have reconceptualized PII as ''PII 2.0,'' a model that categorizes PII as ''identified'' or ''identifiable'' and further recognizes a category of ''non-identifiable'' information.

PII 2.0: Privacy and a New Approach to Personal Information

BY PAUL M. SCHWARTZ AND DANIEL J. SOLOVE

Personally Identifiable Information (PII) is one of the most central concepts in privacy regulation. It defines the scope and boundaries of a large range of privacy statutes and regulations. Numerous federal

Paul M. Schwartz is Professor of Law at the University of California, Berkeley School of Law, and Director of the Berkeley Center for Law & Technology. Daniel J. Solove is the John Marshall Harlan Research Professor of Law at George Washington University Law School. He is also senior policy advisor at Hogan Lovells, and founder of TeachPrivacy, http://teachprivacy.com, a company that helps schools with training and education about privacy issues as well as developing a comprehensive privacy program.

statutes turn on this distinction.[1] Similarly, many state statutes also rely on PII as a jurisdictional trigger.[2] These laws all share the same basic assumption—that in the absence of PII, there is no privacy harm. Thus, privacy regulation focuses on the collection, use, and disclosure of PII, and leaves non-PII unregulated.

Given PII's importance, it is surprising that information privacy law in the United States lacks a uniform definition of the term. In addition, computer science has shown that the very concept of PII is far from straightforward. Increasingly, technologists can take information that appears on its face to be non-identifiable and turn it into identifiable data.

In our view, PII must be reconceptualized if privacy law is to remain effective in the future. Therefore, we have developed a new conception, PII 2.0, which avoids the problems and pitfalls of current approaches. The key to our model is to build two categories of PII, ''identified'' and ''identifiable'' data, and to treat them differently. This approach permits tailored legal protections built around different levels of risk to individuals. This essay is an abridged version of our longer treatment of

---

[1] Examples include the Children's Online Privacy Protection Act, the Gramm-Leach Bliley Act, the HITECH Act, and the Video Privacy Protection Act.

[2] Examples include California's Song-Beverly Credit Card Act, Cal. Civ. Code § § 1747–1748.95, which is available at http://op.bna.com/pl.nsf/r?Open=kjon-8qplsc, as well as numerous state breach notification laws.

this theme—a law review article published by the New York University Law Review.[3]

I. PII's Central Role and Uneasy Status

Given the ubiquity of the concept of PII in privacy law and the important role it plays, the definition of PII is crucial. But instead of defining PII in a coherent and consistent manner, privacy law offers multiple competing definitions, each with significant problems and limitations. There are three predominant approaches to defining PII in various laws and regulations. We will refer to these approaches as (1) the "tautological" approach, (2) the "non-public" approach, and (3) the "specific-types" approach.

A. The Tautological Approach

The tautological approach is an example of a standard, and it defines PII as any information that identifies a person. The Video Privacy Protection Act (VPPA) neatly demonstrates this model.[4] The VPPA, which safeguards the privacy of video sales and rentals, simply defines "personally identifiable information" as "information which identifies a person."[5] For purposes of the statute, information that identifies a person is PII and falls under the statute's jurisdiction once linked to the purchase, request, or obtaining of video material.

The problem with the tautological approach is that it fails to define PII or explain how it is to be singled out. At its core, this approach simply states that PII is PII. As a result, this definition is unhelpful in distinguishing PII from non-PII.

B. The Non-Public Approach

A second approach to defining PII is to focus on non-public information. The non-public approach seeks to define PII by focusing on what it is *not* rather than on what it is. Instead of saying that PII is simply that which identifies a person, the non-public approach draws on concepts regarding information that is publicly accessible and information that is purely statistical. This model would exclude information that falls in these categories from PII, but the relevant legislation does not explore or develop the logic behind this approach.

The Gramm-Leach-Bliley Act (GLBA) epitomizes one aspect of this approach by defining "personally identifiable financial information" as "nonpublic personal information."[6] The statute fails to define "nonpublic," but presumably this term means information not found within the public domain. In an illustration of another aspect of this approach, the Cable Act defines PII as something other than "aggregate data."[7] This statute, which protects the privacy of subscribers to cable services, views PII as excluding "any record of aggregate data which does not identify particular persons."[8] By aggregate data, the Cable Act presumably means purely statistical information that does not identify specific individuals.[9]

The problem with the non-public approach is that it does not map onto whether the information is in fact identifiable. The public or private status of data often does not match up to whether it can identify a person or not. For example, a person's name and address, which clearly identify an individual, nevertheless might be considered public information, as such information is typically listed in telephone books. In many cases, however, individuals have non-public data that they do not want matched to this allegedly public information. Yet, an approach that only protects non-public information as PII might not preclude such combinations.

C. The Specific-Types Approach

The third approach is to list specific types of data that constitute PII. In the context of the specific-types approach, if information falls into an enumerated category, it becomes per se PII by operation of the statute. To illustrate this approach, we will examine Massachusetts's breach notification statute of 2007 (officially titled the Standards for the Protection of Personal Information of the Residents of the Commonwealth) and the federal Children's Online Privacy Protection Act (COPPA) of 1998.

The Massachusetts breach notification statute requires that individuals be notified if a defined set of their personal information is lost or leaked.[10] The Act defines PII as a person's first name and last name, or first initial and last name in combination with either a Social Security number, driver's license number, financial account number, or credit or debit card number.[11]

Second, COPPA, a federal statute, regulates the collection and use of children's information by internet websites or online services.[12] Like the Massachusetts statute, it approaches the question of PII versus non-PII in a typological fashion. COPPA states that personal information is "individually identifiable information about an individual collected online," including first and last name, physical address, Social Security number, telephone number, and e-mail address.[13] This law's definition of PII also includes "any other identifier that the [Federal Trade Commission (FTC)] determines permits the physical or online contacting of a specific individual."[14] In 2000, the FTC issued its COPPA Rule.[15] It added one element to the Act's definition of PII by extending this concept to a "persistent identifier, such as a customer number held in a cookie or a processor se-

---

[3] Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L.Rev. 1814 (2011), *available at* http://ssrn.com/abstract=1909366.

[4] Video Privacy Protection Act of 1988, 18 U.S.C. § 2710 (2006).

[5] *Id.* § 2710(a)(3). The VPPA prohibits "videotape service providers" from knowingly disclosing personal information, such as the titles of items rented or purchased, without the individual's written consent. It defines "videotape service providers" in a technological neutral fashion to permit the law to be extended to DVDs. *Id.* § 2710(a)(4).

[6] Gramm-Leach-Bliley Act of 1999, 15 U.S.C. § 6809(4)(A) (2006).

[7] Cable Communications Policy Act of 1984, 47 U.S.C. § 551(a)(2)(A) (2006).

[8] *Id.*

[9] The number of Comcast customers in Virginia who subscribe to HBO is an example of aggregate data under the Cable Act.

[10] *E.g.*, Standards for the Protection of Personal Information of Residents of the Commonwealth, 201 Mass. Code Regs. § 17.04 (2010).

[11] *Id.* § 17.02.

[12] Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506 (2006).

[13] *Id.* § 6501(8)(A)–(E).

[14] *Id.* § 6501(8)(F).

[15] Children's Online Privacy Protection Act Rule, 16 C.F.R. pt. 312 (2000).

rial number, where such identifier is associated with individually identifiable information.''[16]

An initial problem with the specific-types approach is that it can be quite restrictive in how it defines PII. The Massachusetts statute defines PII to include only a narrow set of data elements: a name plus other elements, such as a Social Security number, a driver's license number, or a financial account number.[17] This list is under-inclusive: there are numerous other kinds of information that, along with a person's name (or independently), would reveal one's identity. Moreover, most individuals would consider such a data breach to be a significant event and one about which they would wish to be informed. The Massachusetts version of the specific-types approach also assumes that the types of data that are identifiable to a person are static—the statute does not cover information that could potentially become personally identifiable. This variant of the specific-types approach is too rigid to adequately protect personal privacy.

As for COPPA, a second example of the specific-types approach, this federal statute has an advantage that the Massachusetts law lacks. COPPA explicitly references FTC rulemaking as a way to expand and adapt its definition of PII.[18] The FTC has indeed acted to expand the definition of PII in the statute; its COPPA rule added one element to the statutory concept of PII, namely, the idea of ''a persistent identifier,'' such as a cookie.[19] However, the FTC's ability to alter the definition of PII is limited by a requirement that information covered by the statute must permit the ''contacting of a specific individual.''[20]

II. The Problems With PII

PII remains a central concept in privacy regulation. It strikes many as common sense that a person's privacy can be harmed only when PII is collected, used, or disclosed. Nonetheless, PII, as currently defined, is a troubled concept for framing privacy regulation. As we contend, the current distinction between PII and non-PII proves difficult to maintain. Indeed, whether information is identifiable to a person will depend upon context and cannot be determined *a priori*.

PII, as currently defined, is a troubled concept for framing privacy regulation.

A. The Anonymity Myth and the IP Address

There is common myth about anonymity on the internet. Many people believe that anonymity exists for most situations when one surfs the web or engages in behavior in cyberspace. The ''anonymity myth,'' as we will call it, is the incorrect assumption that as long as one does not explicitly do something under one's actual

name on the internet, there will be safety from identification. Despite the fact that it appears so easy to act anonymously online, this anonymity offers no more protection than a veil over one's face that can readily be lifted.

At its most basic level, the anonymity myth stems from a mistaken conflation between momentary anonymity with actual untraceability. It is easy to communicate online or surf the web without immediately revealing one's identity, but it is much more difficult to be non-traceable. Whenever one is online, a potential for traceability exists. One threshold issue at the entry to cyberspace that contributes significantly to traceability is the internet protocol (IP) address.

An IP address is a unique identifier that is assigned to every computer connected to the internet. Due to the shift from dial-up to static IP addresses, Internet Service Providers (ISPs) now have logs that link IP addresses with particular computers and, in many cases, eventually to specific users.

The identification of a seemingly anonymous internet user can easily follow from an IP address. Connection to a website normally reveals a user's IP address to the host website, and look-up tools available on the internet permit certain information to be revealed about an IP address. Such details include the host name, geographic coordinates, and a map indicating its general location. With such access to a user's IP address, a third party need only have the user's ISP to match the relevant account information to the IP address assigned to that user's computer in order to personally identify the account holder.

To be sure, IP addresses do not directly identify a particular person. Instead, an IP address is assigned to a specific computer or internet device in order to allow it to access to the internet. Therefore, identification does not follow automatically from access to an IP address alone. For example, a computer may be used by multiple members of a household. Not surprisingly then, some companies have argued that an IP address is non-PII. Yet, this argument is misleading. In the case of the IP address, various other clues can readily be used to identify particular individuals. These clues include analysis of the websites that a person visited during a particular session of web surfing. For example, a family member may check her work webmail and use a unique password to do so. In this fashion, it will be possible to distinguish one member of the family from another. IP addresses can also be readily linked to individuals who post information online.

B. The Re-Identification of Data: Goodbye Non-PII?

Technology is now posing a considerable challenge to the non-PII side of the dichotomy. Computer scientists are finding ever more inventive ways to combine various pieces of non-PII to make them PII. This trend shows up, for example, in some remarkable demonstrations of how supposedly de-identified information can be re-personalized. A major problem with defining some types of information as non-PII is that technology increasingly enables the combination of various pieces of non-PII to produce PII. A further problem with non-PII is the wide availability of so much information about people. This phenomenon of data availability heightens the ability to turn non-PII into PII. This aspect of the re-

---

[16] *Id.*
[17] 201 MASS. CODE REGS. § 17.02 (2010).
[18] 15 U.S.C. § 6501(8)(F) (2006).
[19] 16 C.F.R. § 312.2 (2011).
[20] 15 U.S.C. § 6501(8)(F).

personalization problem stems from a privacy problem that we will call "aggregation."[21] Aggregation involves the combination of various pieces of data.

As we have seen, visitation patterns can permit the use of an IP address to link de-identified data to names and addresses. Additionally, a person who thinks she is anonymous while using certain sites may provide explicitly identifying information, as when completing a purchase. A further example involves a study of Netflix movie rentals by two computer scientists, Arvind Narayanan and Vitaly Shmatikov. The Narayanan-Shmatikov research demonstrated that at least some people in a supposedly anonymous dataset could be identified based on how they rated movies on a publicly available website.[22]

When aggregated, information has a way of producing more information, such that de-identification of data becomes more difficult. Thus, it becomes possible to look for overlap in the data and then to link up different bodies of data.

C. The Problem of Changing Technology and Information-Sharing Practices

A further challenge to current concepts of PII is that technology is constantly changing. Thus, today's non-PII might be tomorrow's PII. New and surprising discoveries are constantly being made about ways of combining data to reveal other data. For example, a recent study by Alessandro Acquisti and Ralph Gross demonstrates that people's Social Security numbers (SSNs) can be predicted based on other pieces of data such as birth date and birth location.[23] As they state, "it is possible to predict, entirely from public data, narrow ranges of values wherein individual SSNs are likely to fall."[24]

In addition to new technological abilities that permit the re-identification of data, another important factor that facilitates re-identification of data is the proliferation of personal information online and in offline record systems. In particular, corporate practices now play an important role in shaping the amount and kinds of information that are available online. Whether information can be re-identified depends on technology and corporate practices that permit the linking of de-identified data with already-identified data. Moreover, as additional pieces of identified data become available, it becomes easier to link them to de-identified data because there are likely to be more data elements in common.

D. The Ability to Identify Depends on Context

In many cases, a determination of whether some data are PII as opposed to non-PII is complex because information does not readily fit into one of these two categories. As noted above, identifiability is a complex concept because of the changing landscape of technology, as well as social and corporate practices. Abstract de-

terminations of whether a given piece of information is PII are insufficient because the ability to identify information is driven by context.

Consider internet search queries that are anonymized. A search query is the information that a person types into a search engine like Google. In the abstract, if anonymized, search queries appear to be non-PII. Yet, it is not possible to make an abstract judgment of whether or not a search query can become PII. It depends upon the nature of the search in which the subject person had been engaged. If the only data are a single search query for something general (such as a search for "poodles") then identifying a specific user might be difficult. But if the user has engaged in a highly specific search, or multiple searches, she becomes more identifiable. At some point, a search allows a person to be readily identifiable.

Thus, the question of whether search queries are PII cannot be answered in the abstract. Trying to classify search queries as PII or non-PII in order to fit them into the binary system of much current privacy regulation is futile. The consequences of search queries will depend upon the context, such as the specific things searched for, as well as what other information is already available about a user. Nonetheless, the distinction between PII and non-PII is almost always made in the abstract in privacy regulation.

III. PII 2.0

The existing definitions of PII have proven problematic. Nonetheless, we reject the idea that privacy law should abandon the concept of PII. If it did so, privacy law would be left without a means for establishing coherent boundaries on necessary regulation. Therefore, we reconceptualize the current standard by introducing PII 2.0.

A. An Explanation of PII 2.0

Our model places information on a continuum that begins with no risk of identification at one end, and ends with identified individuals at the other. We divide this spectrum into three categories, each with its own regulatory regime: under the PII 2.0 model, information can be about an (1) identified, (2) identifiable, or (3) non-identifiable person. Our three categories divide up this spectrum and provide different regimes of regulation for each. Because these categories do not have hard boundaries, we define them in terms of standards.

Information refers to an *identified* person when it singles out a specific individual from others. Put differently, a person has been identified when her identity is ascertained. There is general international agreement about the content of this category, albeit not of the implications of being placed in it. For example, in the United States, the General Accounting Office, Office of Management and Budget, and National Institute of Standards and Technology associate this concept with information that distinguishes or traces a specific individual's identity.[25] In Europe, the Article 29 Group states that a person is identified "when, within a group

---

[21] *See* DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 117–21 (2008) (explaining the mechanics of aggregation).

[22] Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets* (2008 IEEE Symp. on Sec. and Privacy 111, Feb. 5, 2008), *available at* http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf.

[23] Alessandro Acquisti & Ralph Gross, *Predicting Social Security Numbers from Public Data,* 106 PNAS 10975 (2009).

[24] *Id.* at 10975.

[25] ERIKA MCCALLISTER ET AL., NAT'L INST. OF STANDARDS AND TECH., GUIDE TO PROTECTING THE CONFIDENTIALITY OF PERSONALLY IDENTIFIABLE INFORMATION (PII) 2-1 (2010); U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-536, PRIVACY: ALTERNATIVES EXIST FOR ENHANCING PROTECTION OF PERSONALLY IDENTIFIABLE INFORMATION 1 n.1 (2010).

---

of persons, he or she is 'distinguished' from all other members of the group."[26]

Information in the middle of the risk continuum relates to an *identifiable* individual when specific identification, while possible, is not a significantly probable event. In other words, an individual is identifiable when there is some non-remote possibility of future identification. The risk level for such information is low to moderate. Information of this sort should be treated differently from an important subcategory of nominally identifiable information, in which linkage to a specific person has not yet been made, but where such a connection is more likely. As we shall explain, such nominally identifiable data should be treated the same as identified data.

At the other end of the risk continuum, *non-identifiable* information carries only a remote risk of identification. Such data cannot be said to be relatable to a person, taking account of the means reasonably likely to be used for identification. In certain kinds of data sets, for example, the original sample is so large that other information will not enable the identification of individuals. An example would be high-level information about the populations of the United States, China, and Japan, and their relative access to telecommunications.

There are certain instances where identifiable information should be treated like information referring to an identified person. Information that brings a substantial risk of identification of an individual should also be treated as referring to an identified person. In other words, identifiable data should be shifted to the *identified* category when there is a significant probability that a party will make the linkage or linkages necessary to identify a person. This essential subcategory requires assessment of the means likely to be used by parties with current or probable access to the information, as well as the additional data upon which they can draw. This test, like those for the other categories, is a contextual one. It should consider factors such as the lifetime for which information is to be stored, the likelihood of future development of relevant technology, and parties' incentives to link identifiable data to a specific person.

Practical tools also exist for assessing the risk of identification. In fact, computer scientists have developed metrics for assessing the risk of identifiability of information. For example, Khaled El Emam has identified benchmarks for assessing the likelihood that de-identified information can be linked to a specific person—that is, can be made identifiable.[27] The critical axes in El Emam's work concern the "mitigating controls" available to parties in possession of information, and the likely motives and capacity of outsiders who might seek to tie that information to a person. In addition, computer scientists' ongoing work in developing more secure software offers useful lessons. The relevant need is to focus on: (1) the nature of internal and external threats to a data asset, and (2) the effectiveness of possible countermeasures to those threats.[28]

B. PII 2.0 and Fair Information Practices (FIPs)

In our reconceptualized notion of PII, the key is to think about identification in terms of risk level. PII 2.0 conceives of identifiability as a continuum of risk rather than as a simple dichotomy. A clear way to demonstrate the functioning of this new approach is by considering the applicability of FIPs.

The basic toolkit of FIPs includes the following: (1) limits on information use; (2) limits on data collection, also termed data minimization; (3) limits on disclosure of personal information; (4) collection and use only of information that is accurate, relevant, and up-to-date (data quality principle); (5) notice, access, and correction rights for the individual; (6) the creation of processing systems that the concerned individual can know about and understand (transparent processing systems); and (7) security for personal data.[29] When information refers to an *identified* person, all of the FIPs generally should apply.

As for the category of *identifiable*, it is not appropriate to treat such information as fully equivalent to identified. The information does not yet refer to a specific person and may never do so. Nonetheless, some protections are in order because there is a risk of linkage to a specific individual. The question then becomes, which of the FIPs should apply?

Full notice, access, and correction rights should *not* be granted to an affected individual simply because identifiable data about her are processed. For one thing, if the law created such interests, these obligations would decrease rather than increase privacy by requiring that all such data be associated with a specific person. This connection would be necessary in order to allow that individual to exercise her rights of notice, access, and correction. In this fashion, the law would create a vicious circle that could transform identifiable data into identified data. Moreover, limits on information use, data minimization, and restrictions on information disclosure should not be applied across the board to identifiable information. Such limits would be disproportionate to risks from data use and also would cripple socially productive uses of analytics that do not raise significant risks of individual privacy harms.[30]

At the same time, some FIPs should apply to identifiable data. The key FIPs are those that concern data security, transparency, and data quality. Data security refers to the obligation to "protect against unauthorized access to and use, destruction, modification, or disclosure of personal information."[31] Identifiable informa-

---

[26] Article 29 Data Protection Working Party, *Opinion 4/2007 on the Concept of Personal Data* at 12, 01248/07/EN/WP 136 (June 20, 2007).

[27] *See* Khaled El Emam, *Heuristics for De-Identifying Data*, SECURITY & PRIVACY, July/Aug. 2008, at 58; Khaled El Emam, *Risk-Based De-Identification of Health Data*, SECURITY & PRIVACY, May/June 2010, at 64.

[28] *See* MICHAEL HOWARD & STEVE LIPNER, THE SECURITY DEVELOPMENT LIFECYCLE (2006) (discussing techniques for engineers to develop more secure software).

[29] DANIEL J. SOLOVE & PAUL M. SCHWARTZ, PRIVACY LAW FUNDAMENTALS 56 (2011).

[30] At the Article 29 Working Party of the European Union, there recently has been openness to a concept of proportionality in the use of information privacy law. *See* Article 29 Data Protection Working Party, *Opinion 3/2010 on the Principle of Accountability* at 3, 00062/10/EN/WP 173 (July 13, 2010). The question remains as to how successful this concept will be in a system that treats identified and identifiable data as equivalents.

[31] PRIVACY AND SECURITY DESKBOOK 14-3 (Lisa J. Sotto, ed., 2010).

tion should be subject to data security principles. Recall that identifiable data are those for which a specific identification, while possible, is not a significantly probable event. Yet these data, unlike non-identifiable information, might be relatable to a person. Data security for identifiable information, as for identified information, should be commensurate with the nature of the information and the likely risks of disclosure.

As for transparency, this FIP calls for the creation of data processing systems that are open and understandable to affected individuals. Transparency also means that tracking or surveillance should not be done secretly. This FIP is important for identifiable data for two reasons. First, openness about information use allows for improved policies and law. As Louis Brandeis famously stated, "Sunlight is said to be the best of disinfectants; electric light the most efficient policeman."[32]

Second, identifiable information can have great value. Transparency about the collection of identifiable information will serve to heighten awareness about data flows among all parties, both consumers and corporations. It will also improve the position of consumers who have preferences about the collection and further use of their data.

Finally, data quality is a FIP that requires organizations to engage in good practices of information handling. This requirement depends on the purpose for which information is to be processed. In the context of *identified* data, for example, it means that the greater the potential harm to individuals, the more precise that the data and its processing must be. Some decisions matter more than others, however, and the stakes are low when the issue is whether or not one receives a coupon for a dollar discount on a case of seltzer. More precision is required in a data system that decides whether or not one receives a mortgage, and determines the interest rate associated with it. In contexts where the decision to be made about a person based on identified data is more important, or the harm to the person potentially greater, there must be higher requirements for data quality.

In the context of *identifiable* information, data quality also requires good practices of information handling. In particular, it requires that companies pay at-

tention to the handling of identifiable information by third parties. If information is non-identifiable, a company can publicly release it or permit third parties access to it without further obligations.

Identifiable information is capable of identification, even if this risk is not significantly probable. Thus, companies cannot merely release it or allow unmonitored access to it. Depending on the potential harm to individuals and the likely threat model, companies should also be required to use a "track and audit" model for some identifiable information. An example would be information used in health care research. Access to such data should be accompanied by obligations that travel with the information. Companies that handle identifiable information can structure these obligations by associating metadata, or information about information, with data sets.

Thus, one benefit of PII 2.0 is that it tailors FIPs to whether information is identified or identifiable. A further benefit of PII 2.0 is that it creates an incentive for companies to keep information in the least identifiable form. If we abandon PII, or treat identified and identifiable information as equivalents, companies will be less willing to expend resources to keep data in the most de-identifiable state practicable.

Conclusion

PII is a challenging conceptual issue at the heart of any system of regulating privacy in the Information Age. If PII is defined too narrowly, then it will fail to protect privacy in light of modern technologies involving data mining and behavioral marketing. Technology will thus make privacy law irrelevant and obsolete. On the other hand, if PII is defined too broadly, then it could encompass too much information, and threaten to transform privacy law into a cumbersome and unworkable regulation of nearly all information. Privacy law must have coherent boundaries, which adequately protect privacy, and which can be flexible and evolving.

In PII 2.0, flexibility follows from a general association of *different* FIPs with identified or identifiable information. An additional safeguard is provided by treating identifiable information with a substantial risk of being identified as a form of identified information. At this point, the risk of being identified has grown too high. Such an approach prevents tactical attempts to use readily-identifiable data in lieu of identified data in order to avoid regulation and responsibility.

---

[32] Louis D. Brandeis, Other People's Money and How the Bankers Use It 92 (1914).