Reconciling Personal Information in the United States and European Union

Paul M. Schwartz* and Daniel J. Solove**

U.S. and EU privacy law diverge greatly. At the foundational level, they differ in their underlying philosophy: In the United States, privacy law focuses on redressing consumer harm and balancing privacy with efficient commercial transactions. In the European Union, privacy is hailed as a fundamental right that can trump other interests. Even at the threshold level—determining what information is covered by the regulation—the United States and European Union differ significantly. The existence of personal information—commonly referred to as "personally identifiable information" (PII)—triggers the application of privacy law. The U.S. and the European Union define this essential term of privacy law quite differently. The U.S. approach involves multiple and inconsistent definitions of PII that are often particularly narrow. The EU approach defines PII to encompass all information identifiable to a person, a definition that can be quite broad and vague. This divergence is so basic that it threatens the stability of existing policy mechanisms for permitting international data flows.

In this Essay, we propose a way to bridge these differences regarding PII. We contend that a tiered approach to the concept of PII (which we call "PII 2.0") represents a superior way of defining PII compared to the current approaches in the United States and European Union. We also argue that PII 2.0 is consistent with the different underlying philosophies of the U.S. and EU privacy law regimes. Under PII 2.0, all of the Fair Information Practices (FIPs) should apply when data refers to an identified person or when there

Copyright © 2014 California Law Review, Inc. California Law Review, Inc. (CLR) is a California nonprofit corporation. CLR and the authors are solely responsible for the content of their publications.

* Jefferson E. Peyser Professor of Law, University of California, Berkeley, School of Law; Director, Berkeley Center for Law & Technology. For their comments and suggestions, we wish to thank Molly Gavin, Scott Goss, Jesse Koehler, Angel Diaz, and David Vladeck. Research for this paper was supported with funding from Qualcomm Incorporated and summer research grants from UC Berkeley School of Law and George Washington University Law School.

** John Marshall Harlan Professor of Law, George Washington University Law School.

is a significant risk of the data being identified. Only some of the FIPs should apply when data is merely identifiable, and no FIPs should apply when there is a minimal risk that the data is identifiable. We demonstrate how PII 2.0 furthers the goals of both U.S. and EU privacy law and how PII 2.0 is consistent with their different underlying philosophies. PII 2.0 thus advances the process of bridging the current gap between U.S. and EU privacy law.

I.
INTRODUCTION

"Personal data" is a central concept in privacy regulation around the world. This term defines the scope and boundaries of many privacy statutes and regulations. Numerous federal and state statutes in the United States turn on the definition of "personal data."[1] Personal data is also commonly referred to as

1. Examples of federal laws include the Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–6506 (2006); the Gramm-Leach-Bliley Act, 15 U.S.C. §§ 6801–6809 (2006); the Health Information Technology for Economic and Clinical Heath Act of 2009, Pub. L. No. 111-5, 123 Stat. 226 (codified as amended in scattered sections of 42 U.S.C.); and the Video Privacy Protection Act, 18 U.S.C. § 2710 (2012). Examples of state laws include California's Song-Beverly Credit Card Act of 1971, Cal. Civ. Code § 1747 (West 2009) and the numerous state breach notification laws. For an up-to-date listing of the final group of statutes, see *State Security Breach Notification Laws*, NAT'L

"personally identifiable information" (PII), and we will therefore use the terms interchangeably. PII is foundational to any privacy regulatory regime because it serves as a jurisdictional trigger: If there is PII, the laws apply. If it is absent, the privacy regulation in question does not apply.

The concept of PII plays a similar role in the privacy law of the European Union.[2] These laws share the same fundamental assumption—that in the absence of PII, there is no privacy right. For this reason, privacy regulation focuses on the collection, use, and disclosure of PII, and leaves non-PII unregulated. Given PII's importance, it is surprising that it lacks a uniform definition.

In the United States, the law provides multiple explanations of this term. In our previous work, we demonstrated the shortcomings of these PII concepts, which frequently focus on whether the data is actually linked to an identified person.[3] By contrast, the European Union has adopted a single definition that broadly defines PII to encompass all information that is *identifiable* to a person. Even if the data alone cannot be linked to a specific individual, if it is reasonably possible to use the data in combination with other information to identify a person, then the information is PII.

The considerable divergence of the PII definitions in the United States and European Union poses significant difficulties for the harmonization of the two legal systems' privacy regimes. These difficulties matter: the variation in legal definitions of PII raises compliance costs for companies who do business in both areas of the world.[4] Additionally, the differing definitions threaten a status quo built around second-order mechanisms for allowing international data transfers.[5] These negotiated solutions, developed beginning in the late 1990s, are unstable today; the policy mechanisms cannot gloss over the considerable differences in the most basic unit of information privacy law, which is the definition of personal information. Moreover, there is already an increasing number of EU objections to one of these mechanisms, the Safe Harbor, and the divergence of PII definitions raises a further threat to the existing privacy status

---

CONFERENCE OF STATE LEGISLATURES, http://www.ncsl.org/Default.aspx?TabId=13489 (last updated Jan. 21, 2014).

2. For a related argument regarding how constitutional data protection in Germany requires the presence of personal data, see Dieter Grimm, *Der Datenschutz vor einer Neuorientierung*, 12 JURISTENZEITUNG 585, 586 (2013).

3. *See* 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).

4. *See* PONEMON INST. LLC, THE TRUE COST OF COMPLIANCE: A BENCHMARK STUDY OF MULTINATIONAL ORGANIZATIONS 2 (2011), *available at* http://www.tripwire.com/tripwire/assets/File/ponemon/True_Cost_of_Compliance_Report.pdf (summarizing findings about privacy and data protection law compliance for multinational organizations).

5. *See infra* Section II.C.3.

quo. These objections and the divergence matter due to the EU's established role as the "privacy cop to the world."[6]

The two systems' disparate treatment of data in situations where the data is merely *identifiable* but the people to whom the data pertains are not currently *identified* has a significant consequence. It leads to key differences between the systems' PII definitions. In a highly significant swath of U.S. privacy law, this information falls outside privacy regulation.[7] In the European Union, however, this data is fully regulated pursuant to the rigorous protections of the EU Data Protection Directive (Directive).[8] It is also regulated under the more recent EU Proposed General Data Protection Regulation of 2012 (Proposed Regulation).[9] This fundamental incongruity in the U.S. and EU regulatory regimes creates significant confusion and impediments to information flow and use.

In previous work, we focused on the approach to PII in U.S. privacy law and criticized the law's disjointed, inconsistent, and often overly narrow definitions of PII. To make privacy law effective for the future, we 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.[10] This approach permits legal protections tailored to different levels of risk to individuals.

In this Essay, we argue that PII 2.0 can do more than serve as the most workable approach for U.S. privacy law. It can also function well for EU privacy law and help harmonize the significantly divergent approaches between U.S. and EU privacy law. This conclusion may appear surprising; it is also far from apparent from our previous work.

Besides functioning differently, EU and U.S. privacy law have different underlying goals and different structures. As an initial matter, EU law views privacy as a fundamental right, while U.S. law considers it one interest that is balanced against others.[11] It may even be secondary to other concerns, such as freedom of speech.[12] In the European Union, privacy law is viewed in broad terms and expressed in omnibus laws that regulate the public and private

6. David Scheer, *For Your Eyes Only–Europe's New High-Tech Role: Playing Privacy Cop to the World*, WALL ST. J., Oct. 10, 2003, at A1.

7. *See infra* Section II.B.

8. *See infra* Section II.A.1.

9. *See infra* Section II.A.2.

10. Schwartz & Solove, *supra* note 3, at 1877–83. As we will discuss below, as part of PII 2.0's harmonization effort, we leave unchanged the EU category of "sensitive" data.

11. *See* Paul M. Schwartz & Karl-Nikolaus Peifer, *Prosser's* Privacy *and the German Right of Personality*, 98 CALIF. L. REV. 1925, 1953–54 (2010); James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151, 1173–76 (2004).

12. *See, e.g.*, Sorrell v. IMS Health Inc., 131 S. Ct. 2653, 2659 (2011) (striking down a Vermont law adding certain privacy-protection measures because "[s]peech in aid of pharmaceutical marketing . . . is a form of expression protected by the Free Speech Clause of the First Amendment.").

sectors alike.[13] In the United States, privacy law is regulated through narrow sectoral laws that focus on specific industries or specific contexts for the use of personal data.[14] Finally, in the European Union, privacy law forbids personal data processing in the absence of a legal basis.[15] In the United States, however, the general approach is to allow personal data processing unless it causes a legal harm or is otherwise restricted by law.[16] Given these differences, it is no surprise that EU privacy law has a much broader definition of PII than U.S. privacy law.

Attempts to harmonize U.S. and EU privacy law by turning EU privacy law into a U.S.-style approach, or vice versa, are unlikely to succeed. Both the United States and European Union are deeply committed to their respective approaches. While policymakers and scholars have been trying for nearly two decades to bring U.S. and EU privacy law closer together, the Proposed Regulation could push the United States and European Union even further apart. In our view, PII 2.0 can serve as a foundational step in overcoming the differences between U.S. and EU privacy law. In this Essay, we set forth the argument for a tiered approach to the concept of PII as a way to bridge trans-Atlantic privacy differences.

II.
DEFINING PII ON BOTH SIDES OF THE ATLANTIC

A comparative focus is necessary to understand the modern landscape of information privacy law. Legal forces outside the United States have significantly shaped the governance of information privacy. In particular, the European Union has played a major role in international decisions that have developed and shaped this area of law. This role has been bolstered by EU laws granting member states the authority to block data transfers to third-party nations, including the United States.

A. The European Union: From the Directive to the Proposed Regulation

In the European Union, the current framework for defining personal information includes both the Directive, which was enacted in 1995,[17] and the Proposed Regulation, which was released in 2012[18]—the final form of which

13. *See* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 1110 (4th ed. 2011).

14. *See* DAVID H. FLAHERTY, PROTECTING PRIVACY IN SURVEILLANCE SOCIETIES 404–05 (1992).

15. Directive 95/46/EC, of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, art. 7, 1995 O.J. (L 281) 31–32 [hereinafter Data Protection Directive].

16. Paul M. Schwartz, *Preemption and Privacy*, 118 YALE L.J. 902, 913 (2009).

17. Data Protection Directive, *supra* note 15.

18. *Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement*

EU institutions are currently debating. The Directive, which is in force, sets out common rules for data protection in EU member states and requires these countries to enact legislation that follows the Directive's standards. Although the Directive employs the term "personal data," this term serves the same function as PII, and in this Essay, we treat the legal terms "personal data" and "PII" as functional equivalents.

Under both the Directive and Proposed Regulation, the EU takes a broad approach to defining PII. The definition turns on whether a natural person is capable, directly or indirectly, of identification through a linkage or some other reference to available data. In the European Union, information that is identified or identifiable receives an equivalent level of legal protection.

1. The EU Data Protection Directive of 1995

The Directive uses the term "personal data" and defines it as "information relating to an identified or identifiable natural person."[19] The Directive does not explicitly define "identified." Under an EU directive, the law of member states then becomes determinative. Among EU member states that have traditionally taken a leading role in information privacy law, a person falls in the "identified" category if a party can use information relating to her to determine her specific identity.[20] In analyzing the term under German law, for example, Ulrich Dammann states, "A person is identified when it is clear that the data relate to the person and not to another."[21] Concerning the law of the United Kingdom, Rosemary Jay writes, "A person becomes *identified* where there is sufficient information either to contact him or to recognise him by picking him out in some way from others and know who he/she is."[22] In France, the national data protection commission has simply explained that a person is identified "if, for example, his name appears in a file."[23]

The Directive is more specific regarding its definition of "identifiable." It explains that an "identifiable" person is "one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."[24] As additional definitional assistance, the Directive in its Recital 26 explains that in determining whether a person is identifiable,

---

of such Data (General Data Protection Regulation), COM (2012) 11 final (Jan. 25, 2012) [hereinafter *Proposed Regulation*].

19. Data Protection Directive, *supra* note 15, at art. 2(a).

20. On the different leadership roles among national and corporate actors in the development of European privacy law, see ABRAHAM L. NEWMAN, PROTECTORS OF PRIVACY 76 (2008).

21. Ulrich Dammann, *§ 3 Weitere Begriffsbestimmungen*, *in* BUNDESDATENSCHUTZGESETZ 297, 310 (Spiros Simitis ed., 7th ed. 2011).

22. ROSEMARY JAY, DATA PROTECTION LAW AND PRACTICE 172 (4th ed. 2012).

23. *Qu'est-ce qu'une donnée personnelle?,* COMMISSION NATIONAL DE L'INFORMATIQUE ET DES LIBERTÉS, http://www.cil.cnrs.fr/CIL/spip.php?rubrique299 (last visited Mar. 23, 2014).

24. Data Protection Directive, *supra* note 15, at art. 2(a).

"account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person."[25] This approach follows a longstanding paradigm in German federal data protection law.[26]

Both identified and identifiable information fall squarely within the scope of EU data privacy law, and they are treated in the same fashion. The Directive specifies obligations on the "data controller," rights for the "data subject," and robust protections for personal data. Before turning to these rights and duties, it should be clarified that as a matter of terminology, EU data privacy law refers to the entity that collects and uses personal data as the "data controller" and the individual whose data is involved as the "data subject."[27] The duties of the data controller and the rights of the data subject are the same for both identified and identifiable information. The crossing of the threshold for either category functions as an "on" switch for the application of EU data protection law.[28]

Once information qualifies as identified or identifiable, it falls under the data protection regime. At that moment, a full suite of obligations and protections is triggered. From the U.S. perspective, the resulting EU legal regime is formidable both in terms of the protection granted to the affected individual—the data subject—and the duties placed on the party who processes the personal information. The general EU rule is that the collection and processing of personal data must be for "specified, explicit and legitimate purposes."[29] These purposes may be ones to which the personal data subject has consented, purposes necessary to protect the data subject's vital interests, purposes in the public interest, or purposes in the legitimate interests of the data controller—unless they interfere with a data subject's fundamental right to privacy.[30]

The Directive also provides data subjects with a right to control the use of their personal data. Data subjects must be informed about the entities that collect their personal information, how it will be used, and to whom it will be transferred.[31] Under the Directive, data subjects also have a right to access their personal data and to correct inaccurate information in their records.[32] The Directive requires that data subjects provide affirmative consent before their

25. *Id.* at pmbl. ¶ 26.

26. *See* Bundesdatenschutzgesetz [BDSG] [Federal Data Protection Act], Jan. 14, 2003, BGBL. I at 66, last amended Aug. 14, 2009, BGBL. I at 2814 (Ger.); Dritter Abschnitt, *§ 33 Benachrichtigung des Betroffenen*, *in* BUNDESDATENSCHUTZGESETZ 1152, 1159 marginal no. 22 (Spiros Simitis ed., 6th ed. 2006).

27. Data Protection Directive, *supra* note 15, at art. 2(a), (d).

28. Once information qualifies as "personal data" under Article 2 of the Directive, the full set of obligations and rights associated with this term become applicable. *See, e.g.*, *id*. at art. 6 ("Principles Relating to Data Quality"), art. 7 ("Criteria for Making Data Processing Legitimate").

29. *Id.* at art. 6(b).

30. *Id.* at art. 7.

31. *Id.* at art. 10.

32. *Id.*

personal data is processed, used, or disclosed.[33] Consent must be unambiguous and freely given. Data subjects have a right to object to a data controller's use of personal data.[34] Data subjects also have a right not to be subject to certain decisions made solely based on the automated processing of data.[35] They are to be informed of the logic used in the automatic processing of that data.[36]

As for data controllers, the Directive imposes a number of obligations on them beyond those that follow from the rights of data subjects. To begin with, data controllers may not process personal information collected for one purpose in ways incompatible with that purpose.[37] Data must be kept accurate and current.[38] Data controllers cannot keep personal information for longer than necessary to accomplish the purposes for which it was collected.[39] Data must be kept secure.[40]

Beyond these general obligations, the Directive also mandates additional protections for certain categories of personal data, or before certain actions may be taken with personal data. This special category concerns "sensitive data," which includes data about "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life."[41] Data controllers must notify supervisory authorities before engaging in many kinds of data processing.[42] Finally, and in a step that has heightened its international significance, the Directive restricts the transfer of personal data to other countries. Personal data may be transferred only to countries with an "adequate level of protection" of privacy.[43]

In sum, in the European Union, information that is identified or identifiable falls under the definition of "personal data." This classification triggers a wide range of obligations, rights, and protections. As we will now see, the Proposed Regulation also treats identified and identifiable as equivalent s. Its new term of art, however, is "indirectly identified" rather than the Directive's term, "identifiable."[44]

---

33. *Id.* at art. 7(a).
34. *Id.* at art. 14.
35. *Id.* at art. 15.
36. *Id.* at art. 12(a).
37. *Id.* at art. 6(b).
38. *Id.* at art. 6(d).
39. *Id.* at art. 6(e).
40. *Id.* at art. 17.
41. *Id.* at art. 8.
42. *Id.* at art. 18. There is wide divergence among EU Member States regarding the implementation of the Article 18 broadly defined obligation to notify the supervisory authority. France and Belgium have traditionally made the greatest use of this requirement. For a classic account of the highly bureaucratic nature of this practice in France in the 1980s, see FLAHERTY, *supra* note 14, at 165–74. On the Belgian requirements, see CHRISTOPHER KUNER, EUROPEAN DATA PROTECTION LAW 252 (2d ed. 2007).
43. Data Protection Directive, *supra* note 15, at art. 25.
44. *See Proposed Regulation*, *supra* note 18, at art. 4(1).

2. The EU Proposed General Data Protection Regulation of 2012

The European Union is now in the process of replacing the Directive with the Proposed Regulation. In January 2012, the European Commission released a draft version of this document, its Proposed General Data Protection Regulation.[45] This development marks an important policy shift. In EU law, the contrast is clear between a directive and a regulation. While a directive requires Member States to pass harmonizing legislation that "transposes" its standards into national law, a regulation establishes directly enforceable standards. As Christopher Kuner explains, "[A] regulation leads to a greater degree of harmonization, since it immediately becomes part of a national legal system, without the need for adoption of separate national legislation; has legal effect independent of national law; and overrides contrary national laws."[46] Due to its directly binding effect, the Proposed Regulation, if finally approved, will be even more important than the Directive. Moreover, it would assume this importance from its first day of enactment because there would be no wait for the enactment of harmonized national legislation, which can take several years in the case of a directive.[47]

The Proposed Regulation generally builds on the approach of the Directive, but contains some notable changes. Instead of a concept of "identified" or "identifiable," it first defines personal data as "any information relating to a data subject."[48] The nature of the "relating to" requirement is further specified in the definition of "data subject." The Proposed Regulation states that a data subject is a person who "can be identified, directly or indirectly, by means reasonably likely to be used."[49] Thus, the Proposed Regulation shifts from the Directive's notion of identified or identifiable to a concept of direct or indirect identification.

At the same time, however, there is important continuity in the concept of "means reasonably likely to be used." The ultimate test regarding "identifiability" (Directive) or indirect identification (Proposed Regulation) is the same. An analysis must consider "all the means likely reasonably to be used

45. *See generally id.* For an introduction to the Proposed Regulation, see Paul M. Schwartz, *The EU-U.S. Privacy Collision: A Turn to Institutions and Procedures*, 126 HARV. L. REV. 1966, 1992–2000 (2013). *See also* Jacob M. Victor, Comment, *The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy*, 123 YALE L.J. 513 (2013).

46. Christopher Kuner, *The European Commission's Proposed Data Protection Regulation: A Copernican Revolution in European Data Protection Law*, 11 PRIVACY & SECURITY L. REP. 215, 217 (2012).

47. For example, the Directive required all EU Members States to have harmonized their legislation by 1998. Christoph Klug, *Directive 95/46/EC – Data Protection Directive*, *in* CONCISE EUROPEAN IT LAW 151 (Alfred Büllesbach et al. eds., 2d ed. 2010). There were notable delays, however, in this process with proceedings initiated in 1999 against France, Germany, Ireland, Luxembourg, and the Netherlands before the European Court of Justice. *Id.* at 153. At present, all Member States have fully implemented the Directive. *Id.* at 152-53.

48. *Proposed Regulation*, *supra* note 18, at art. 4(2).

49. *Id.* at art. 4(1).

either by the controller or by any other person to identify" the individual.[50] The repetition of the language in both documents indicates that when determining whether personal data exists in the European Union, one must consider the likelihood that certain steps, such as combining bits of scattered data or re-identifying nonpersonal information, will be taken.

The Proposed Regulation provides additional examples of the kinds of linkages that tie information, whether directly or indirectly, to a person. The new examples refer to "location data," "online identifier[s]," and "genetic" identity.[51] The impact of these additional categories is to modernize and expand the sweep of the 1995 Directive.[52]

The Proposed Regulation also contains helpful indications of the need for flexibility in deciding when personal information does or does not exist. For example, its Recital 24 provides important limitations on the Proposed Regulation's concept of indirect identification.[53] Recital 24 initially notes that the use of "online services" may lead individuals to "be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet Protocol addresses or cookie identifiers."[54] Such "information received by the services" can lead to identification of individuals by creating profiles and in other ways.[55] At this point, Recital 24 offers its specific language concerning flexibility. It observes that these kinds of associations do not invariably create identifiable information. Recital 24 states: "[I]dentification numbers, location data, online identifiers or other specific factors as such need not necessarily be considered as personal data in all circumstances."[56] This language indicates the potential under the Proposed Regulation for a tailored, context-specific analysis for deciding whether or not personal data is present.

In summary, an identified person in the European Union is one that can be singled out, whether directly or indirectly, through a linkage to information that references her or him. In a fashion that is consistent with the Directive's approach, the Proposed Regulation offers a broad approach to defining personal information. The critical analysis in the European Union remains focused on

---

50. *Id.* at pmbl. ¶ 23; Data Protection Directive, *supra* note 15, at pmbl. ¶ 26.

51. *Proposed Regulation*, *supra* note 18, at art. 4(1).

52. The Proposed Regulation's key language comes in its definition of "data subject." This term refers to the individual whose personal data is processed and who can be identified. The relevant language at Article 4 is worth citing. A data subject is:

> "[A]n identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person."

Id. As this language indicates, the Proposed Regulation drops the Directive's language about "identifiable," but retains its idea of indirect identification.

53. *See id.* at pmbl. ¶ 24.

54. *Id.*

55. *Id.*

56. *Id.*

whether a natural person is capable of identification, based on an analysis of all means likely to be used and by reference to available data. Finally, Recital 24 of the Proposed Regulation points to the use of flexibility in the analysis of when personal information is and is not present.

The breadth of the EU approach has both benefits and drawbacks. The primary benefit is that hardly anything escapes EU privacy regulation. There are few gaps and inconsistencies under the EU approach, a stark contrast to the U.S. approach where such gaps and inconsistencies are legion. But there is also a primary drawback to the EU approach. Under both the Directive and Proposed Regulation, information is treated as the same whether it refers to an identified individual, or one who can be "indirectly identified"—that is, someone who the Directive terms "identifiable." All these terms constitute personal data, and their presence activates the "on" switch for triggering a full suite of obligations and protections. Yet, a broad continuum of identifiable information exists, and it includes different types of anonymous or pseudonymous information. Moreover, different levels of effort are required to identify information, and various risks are associated with the possible identification of data. Placing all such data into the same conceptual category as "data that currently relate to an identified person" lacks nuance. It also risks activating burdensome regulations for data-processing entities that are incommensurate with actual risks to the privacy of individuals.

B. The United States: A Lack of a Uniform Standard

Instead of defining personal information in a coherent and consistent manner, privacy law in the United States offers multiple competing definitions. As an initial matter, the law in the United States at times drops any distinction between "identifiable" and "identified." This point is illustrated by the U.S. conception of "personally identifiable information," a common term for "personal data." This term sweeps in both identified and identifiable data and thereby elides any differences that may exist between them. Several statutes and regulations in the United States adopt this term.[57] The Google Ngram Viewer also demonstrates that "personally identifiable information" has become an increasingly popular term since 1992.[58] Drawing on Google's ambitious digital library project, this product allows graphical representation of the popularity of words and terms in English and other languages.[59] The chart shows a steep increase in the use of the term in English beginning in that year.[60]

57. *See, e.g.*, Video Privacy Protection Act, 18 U.S.C. § 2710(a)(3) (2012).

58. GOOGLE BOOKS NGRAM VIEWER, https://books.google.com/ngrams (type "personally identifiable information" into the text box, then press enter).

59. *Id.*

60. *Id.*

Neither federal nor state law agree on a single term that identifies the basic category of personal information. We have already discussed the term "personally identifiable information," but U.S. law is far from settled on this nomenclature to identify the underlying concept to which it refers. For example, the Children's Online Privacy Protection Act defines "personal information" as "individually identifiable information about an individual."[61] In California, the Song-Beverly Act uses the term "personal *identification* information" and defines it as information concerning a credit cardholder "other than information set forth on the credit card, and including, but not limited to, the cardholder's address and telephone number."[62]

More generally, the moment at which information becomes identifiable enough to fall within the scope of a particular law relies on how each information privacy statute specifically defines its particular concept of personal information. Thus, while the Children's Online Privacy Protection Act (COPPA) defines "personal information" as "individually identifiable information about an individual,"[63] identifiability under it depends on further analysis of the statute as well as recourse to applicable Federal Trade Commission (FTC) regulations.

In U.S. law, there are three predominant approaches to defining personal information. These are (1) the "tautological" approach; (2) the "nonpublic" approach; and (3) the "specific-types" approach.[64] The tautological approach is an example of a standard, or an open-ended decision-making tool.[65] Under the tautological approach, U.S. privacy law simply defines "personal" as meaning any information that identifies a person. The Video Privacy Protection Act of 1988 (VPPA) neatly demonstrates this model.[66] The VPPA, which safeguards the privacy of video sales and rentals, defines "personally identifiable information" as "information which identifies a person."[67] For purposes of the statute, information that identifies a person becomes "personal identifiable information" and falls under the statute's jurisdiction if tied to the purchase, request, or obtaining of video material.

---

61. 16 C.F.R. § 312.2 (2013).

62. Song-Beverly Credit Card Act of 1971, Cal. Civ. Code § 1748.6 (West 2009) (emphasis added).

63. 16 C.F.R. § 312.2 (2013).

64. *See* Schwartz & Solove, *supra* note 3, at 1828–36.

65. The classic example of a standard would be an instruction to drive at a reasonable speed, or, in the law of negligence, to take the precautions of a reasonable person. For a discussion of the distinction between rules and standards, see Carol M. Rose, *Crystals and Mud in Property Law*, 40 STAN. L. REV. 577, 592–93 (1988), and Kathleen M. Sullivan, *The Supreme Court 1991 Term — Foreword: The Justices of Rules and Standards*, 106 HARV. L. REV. 22, 57–59 (1992).

66. *See* 18 U.S.C. § 2710 (2012).

67. *Id.* § 2710(a)–(b). 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 technologically neutral fashion to permit the law to be extended to DVDs. *Id.* § 2710(a)(4).

A second model focuses on nonpublic information. The nonpublic approach seeks to define personal information by focusing on what it is *not* rather than what it is. The nonpublic approach excludes from its protected scope any information that is publicly accessible or that is purely statistical. The relevant legislation does not explore or develop the logic behind this approach, but rather simply concludes that information falling in these categories is not personal information.

The Gramm-Leach-Bliley Act of 1999 (GLBA) epitomizes this approach by defining "personally identifiable financial information" as "nonpublic personal information."[68] The statute fails to define "nonpublic," but presumably this term means information not found within the public domain. The FTC Regulations to the GLBA explain this term in more detail, but they leave confusion as to whether some publicly accessible information may be classified as "nonpublic" for purposes of the statute.[69] The applicable regulations sweep in "any information" that a consumer provides on a financial application, which seems to relax the core statutory concept of "nonpublic."[70]

In an illustration of another aspect of the "nonpublic" approach, the Cable Communications Policy Act of 1984 defines PII as something other than "aggregate data."[71] 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."[72] By aggregate data, the Cable Act probably means purely statistical information that does not identify specific individuals.[73]

The third approach of U.S. privacy law is to list specific types of data that constitute personal information. In the context of the specific-types approach, if information falls into an enumerated category, it becomes per se personal information under the statute. State data breach notification laws take this approach. These statutes, which forty-seven states have now enacted, require a business to notify affected individuals should an information security breach occur.[74]

The typical trigger for these laws turns on the "unauthorized acquisition" or "unauthorized access" of unencrypted personal information.[75] These laws then generally define personal information through the specific-types approach. As an illustration, the Massachusetts breach notification statute requires that

68. 15 U.S.C. § 6809(4)(A) (2012).
69. *See* 16 C.F.R. § 313.3(n) (2013).
70. *See id.* § 313.3(n)(1), (o)(1).
71. *See* 47 U.S.C. § 551(a)(2)(A) (2012).
72. *Id.*
73. The number of Comcast customers in Virginia who subscribe to HBO is an example of aggregate data under the Cable Act.
74. *See* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, PRIVACY LAW FUNDAMENTALS 176–78 (2013).
75. *See id.* at 17–74.

individuals be notified if a specific set of their personal information is lost or leaked.[76] The Massachusetts law defines personal information as a person's first name and last name, or first initial and last name in combination with a social security number, driver's license number, financial account number, or credit or debit card number.[77] PII is present only when a last name and first name, or last name and first initial, are combined with at least one of the enumerated types of data.

To complicate the lives of lawyers handling multistate data breaches, other state statutes contain different lists of enumerated types of information that constitute PII.[78] As Lisa Sotto notes in her privacy law treatise, "Many states have varied the definition of personal information" to include not only the elements listed above, as found in the Massachusetts law, but also "any number of other potentially identifiable data elements."[79] For example, a 2013 amendment to the California breach notification law expands the statute's definition of personal information to include "a user name or email address, in combination with a password or security question and answer that would permit access to an online account."[80]

At the same time, however, these state laws all share something essential: the presence of personal information is necessary to trigger data breach notification. The laws also generally, although not unanimously, agree as to when PII is present. As Sotto writes, these laws impose a duty to notify only where the data combines "a state's resident's first name, or first initial and last name" in combination with certain other enumerated elements.[81] The flaw of the majority of these statutes is clear: certain information beyond names and initials is readily capable of identifying a specific individual. State breach notification statutes should cover a breach of such information. Only a few states have a trigger in their data security breach notification laws other than first name or initial and last name. As examples of the minority approach, Georgia, Maine, and Oregon have general "savings clauses" that extend protection to data elements even when they are not connected to a person's name if the information would be sufficient to permit identity theft.[82] These states are leading the way for better, next-generation data breach notification laws.[83]

---

76. MASS. GEN. LAWS ANN. ch. 93H, § 3 (West 2007).

77. *Id.* § 1.

78. SOLOVE & SCHWARTZ, *supra* note 74, at 176–78.

79. LISA J. SOTTO, PRIVACY AND DATA SECURITY LAW DESKBOOK § 15.02[B] (2013).

80. S. 46, 2013 Leg., Reg. Sess. (Cal. 2013) (amending CAL. CIV. CODE § 1798.82(h) (West 2014)).

81. SOTTO, *supra* note 79, at § 15.02[B].

82. GA. CODE ANN. § 10-1-911(6) (West 2003); ME. REV. STAT. ANN. tit. 10, § 1347(6) (2014); OR. REV. STAT. § 646A.602(11) (2011).

83. For a chart exploring and categorizing the various PII definitions in different state data security breach notification laws, see SOLOVE & SCHWARTZ, *supra* note 74, at 176–78.

One can also point to a broader flaw in the approach of U.S. law to PII. As a general rule, PII in the United States is largely limited to instances where data refers to an *identified* individual. The exception that proves the rule is the FTC's new regulation to COPPA: its definition of personal information includes a "persistent identifier that can be used to recognize a user over time and across different Web sites or online services."[84] This rule, driven by the FTC's strict policy concerns for protecting children on the Internet, shifts identifiable data into the category of identified.[85] By contrast, the more typical U.S. approach is represented by the many state breach notification laws. In these statutes, personal information is limited to identified data: namely, a first name or initial and a last name, plus information from a specific list of the kinds of data that will make identification certain.

Similarly, the FTC regulations to the GLBA focus on identified data. Under the broad definition given in the regulations, "nonpublic personal information" includes a person's name plus such information as a social security number, driver's license number, and credit or debit card number.[86] This definition clearly contrasts with conceptions of PII in the European Union, where data protection law extends expansively to any data that is identifiable (i.e., that could possibly be linked to an individual).

C. Personal Information: A Problem on Both Sides of the Atlantic

Current approaches to defining PII in the United States and in the European Union are all flawed in significant ways. Stated succinctly, we find the EU approach to be too expansionist and the U.S. approach too reductionist, with problematic consequences flowing from both techniques. Moreover, the divergence between these approaches raises the threat of destabilizing the privacy status quo between the United States and European Union. The stakes in this area of law are high because the trans-Atlantic free flow of data depends on coordination between the two legal systems.

If PII or personal data were a small dimension of privacy regulation, such problems might be isolated and worked around. But the definition of PII is a foundational issue that implicates the scope of privacy regulation. Before even considering differences in *how* data is protected in the United States and European Union, we must address differences in *what* data is protected under these two privacy regimes. This state of disarray points to the critical importance of revising the definition of PII in the United States and European Union alike.

---

84. Children's Online Privacy Protection Rule, 78 Fed. Reg. 3972 (Jan. 17, 2013) (to be codified at 16 C.F.R. § 312.2).

85. *See id.*

86. *See* 16 C.F.R. § 313.3(n)(1), (o) (2013).

1. Evaluating the EU Approach

The benefit of the EU approach to personal information is that it recognizes the expanding ability of technology to re-identify information and to link scattered crumbs of information to a specific individual. The instruction to take "account . . . of all the means likely reasonably to be used" to identify a person results in a flexible, context-based standard.[87] As noted, moreover, the Proposed Regulation states, "identification numbers, location data, online identifiers or other specific factors as such need not necessarily be considered as personal data in all circumstances."[88] Here, too, the Regulation indicates that an evaluation should consider whether possible steps in combination or re-identification of data are likely to be taken.

Despite this promise, the EU's definition of personal information risks sweeping too broadly. Much depends on judgments about open-ended factors, such as "the means likely reasonably to be used,"[89] and the tests developed within the European Union for evaluating such terms are not reassuring. White Papers of the Article 29 Working Party (Working Party) illustrate this point.[90] The Working Party is an important group of EU national data protection commissioners.[91] Under the Directive, it has an advisory role in contributing to "the uniform application" of national privacy law.[92] As such, its opinions on issues such as the definition of personal data provide an important window into EU privacy law. But its approach to defining personal data proves to be flawed.

In its 2007 opinion "On the Concept of Personal Data," the Working Party presents a number of overarching principles for deciding when personal information is present, as well as some illustrations that reveal problematic aspects of its chosen approach.[93] At the same time, the core insight of the Working Party is sound: in looking at whether information is personal data, the analysis must be a "dynamic one" that considers, among other factors, the "state of the art" of the relevant technology and how it is likely to advance over the information's life cycle.[94]

While there is merit in this dynamic analysis, the 2007 Opinion also relies upon the idea that the Directive contains a "broad notion of personal data."[95] This document's wide conception, in turn, is said to further the Directive's objective, which is to protect "the fundamental rights and freedoms" of

87. Data Protection Directive, *supra* note 15, at pmbl. ¶ 26.

88. *Proposed Regulation*, *supra* note 18, at pmbl. ¶ 24.

89. *Id.* at pmbl. ¶ 23.

90. *See* Article 29 Data Protection Working Party, Opinion 4/2007 on the Concept of Personal Data, 01248/07/EN/WP 136 (June 20, 2007).

91. For background on this EU institution, see NEWMAN, *supra* note 20, at 75–76.

92. Data Protection Directive, *supra* note 15, at art. 30(1)(a).

93. *See generally* Article 29 Data Protection Working Party, *supra* note 90.

94. *Id.* at 15. The idea of the informational life cycle is that data can exist in distinct periods and conditions from its first collection to its disposal or destruction.

95. *Id.* at 4.

individuals.[96] As a consequence, the Working Party warns against anyone acting to "unduly restrict the interpretation of the definition of personal data."[97] It then points to the need for a dynamic analysis of when personal data is present. While this language points to a useful approach to assessing the presence or absence of "personal data"—and, in particular, to the need to consider the latest developments in computer science and the data's likely life cycle—the Working Party's own interpretation of these concepts is far from unproblematic.

We will first outline the two key flaws in the 2007 Opinion and then analyze each in turn. First, the Working Party redefines "personal data" as involving decision making based on specific characteristics of the person.[98] Yet, this is a different issue from that of identifiability. Second, the Working Party views information as per se identifiable if the ultimate purpose of the data controller is to identify *some* of the parties in the database.[99] This approach is far different than one that estimates the likelihood of identification. Moreover, it moves an even greater analytic distance from consideration of the risk to a specific person—that is, from a harm analysis. Its focus is on the moment of collection of data and the processing purpose.

The Working Party's examination of web tracking illustrates the first point. The Working Party concludes that a unique identifier assigned to a computer on the Web creates personal information because "web traffic surveillance tools make it easy to identify the behaviour of a machine and, behind the machine, that of its user."[100] In language worth quoting at length, the Working Party states:

> Without even enquiring about the name and address of the individual it is possible to categorise this person on the basis of socio-economic, psychological, philosophical or other criteria and attribute certain decisions to him or her since the individual's contact point (a computer) no longer necessarily requires the disclosure of his or her identity in the narrow sense.[101]

This analysis moves information into the "identifiable" category because of two combined factors: an identifier's link to a specific computer, and the data processor's computer-driven decision making about personal characteristics, such as "socio-economic, psychological, philosophical or other criteria."[102]

This focus of the Working Party seems influenced by a longstanding concern in the European Union regarding decision making about a person based on "automatic means." The underlying worry is about computerized

96. Data Protection Directive, *supra* note 15, at art. 1(1).
97. Article 29 Data Protection Working Party, *supra* note 90, at 5.
98. *See id.* at 12.
99. *See id.* at 13.
100. *Id.* at 14.
101. *Id.*
102. *Id.*

judgments without final human intervention in the process.[103] As previously noted, the Directive provides protection against such decision making.[104] The Proposed Regulation follows this path: it requires limits on automated decision making and ties its concern to current concerns about profiling. Its Article 20 states:

> Every natural person shall have the right not to be subject to a measure . . . which is based solely on automated processing intended to evaluate certain personal aspects relating to this natural person or to analyse or predict in particular the natural person's performance at work, economic situation, location, health, personal preferences, reliability or behaviour.[105]

Here is an important distinction with U.S. information privacy law, which does not generally single out "automated" decision making for special regulation.

More broadly, the Working Party is concerned about how computerization allows individuals to be characterized based on their data trails and then placed in categories, which are then associated by computers with behavioral predictions. This, in short, is the world of Big Data, in which computers driven by algorithms look for correlations amidst a sea of information.[106]

In our view, however, the necessary analysis should differ from that of the Working Party. Web tracking, through means such as the placing of alphanumerical codes on an individual's computer, raises a host of complex issues and, in some cases, significant risks of privacy violations. For example, while contemporary advertising networks may not know people's names, identification of specific individuals is nonetheless possible in many cases.[107] In certain circumstances, enough pieces of information can be associated with a person through her use of a computer to make the process of identification a genuine possibility.[108] At other times, this identification will not be likely to occur, which means there is no use of personal information.

Privacy harms require data use or disclosure pertaining to a specific individual who is identified or reasonably identifiable. In our view, identified information is present when a person's identity has been ascertained, or when there is substantial risk of identification of a specific person by a party who is likely to obtain that information. Targeted marketing that categorizes persons

---

103. *Id.* at 5 ("The processing of personal data by non-automatic means is only included within the scope of the Directive where the data form part of a filing system or are intended to form part of such system (Article 3).").

104. *Supra* Section II.A.

105. *Proposed Regulation*, *supra* note 18, at art. 20(1).

106. *See generally* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA (2013).

107. *See* Emily Steel & Julia Angwin, *On the Web's Cutting Edge, Anonymity in Name Only*, WALL ST. J., Aug. 4, 2010, http://online.wsj.com/news/articles/SB10001424052748703294904575385532109190198.

108. *See* Michael Barbaro & Tom Zeller Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006, http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0.

on the basis of socioeconomic and other criteria can raise issues about consumer protection and discrimination toward or against certain groups.[109] Yet, unless this gathering of information creates data that is reasonably capable of being linked to a specific person, it does not create identified information. Depending on the precise safeguards that the web-tracking company takes, this information may only be identifiable, or even nonpersonal data.

As a second problematic element in the 2007 Opinion, the Working Party views any information that is stored as per se identifiable if the ultimate purpose of the data controller is to identify *some* of the parties in the database.[110] The Opinion's specific examples concern video surveillance, dynamic IP addresses, and the recording of graffiti tags by a transportation company.[111] The problem is that the Working Party's approach confuses collection and stated purpose with identifiability. As a result, it considers identifiable information as present even in circumstances when most or even all of the information in question is never identified.[112] We can elaborate on this analysis by exploring how the Working Party reaches its conclusion regarding how the identification of *some* parties makes *all* of the information in question identifiable.

The Working Party's logic is straightforward. It frontloads the analysis in a fashion that turns the *collection* of information and the overall stated *purpose* into the decisive events for analysis of whether personal data are present. It argues that if "the purpose of the processing implies the identification of individuals, it can be assumed that the controller or any other person involved have or will have the means 'likely reasonably to be used' to identify the data subject."[113] In each case, so long as the ultimate intention is to link some of these data to individuals, all of the information—including that never tied to any person—is treated as personal data.

The model of the Working Party transforms all of the information from the moment of its collection into identifiable data, which receives the same status as identified information in the European Union. This approach is further illustrated by a final example in the Working Party's opinion on personal data. In it, the Working Party considers "key-coded data," which is typically used in

109. *See, e.g.*, Community Reinvestment Act of 1977, Pub. L. No. 95-128, §§ 801-06, 91 Stat. 1147 (1977) (codified as amended at 12 U.S.C. §§ 2901–2908) (preventing lenders from discriminatory credit practices against persons residing in low-income neighborhoods, a practice known as redlining). *See generally* TIMOTHY P. GLYNN ET AL., EMPLOYMENT LAW 515–43 (2007) (discussing antidiscrimination law).

110. *See* Article 29 Data Protection Working Party, *supra* note 90, at 16.

111. *Id.* at 16–17.

112. *Id.* at 16 (noting, for example, in its video surveillance illustration that "[a]s the purpose of video surveillance is . . . to identify . . . persons . . . *where such identification is deemed necessary* by the controller, the whole application as such has to be considered as processing data about identifiable persons, even *if some persons recorded are not identifiable in practice*" (emphasis added)).

113. *Id.*

clinical trials with medicines.[114] In such clinical trials, a key permits identification of individual patients.[115] Such identification is needed if, for example, medicines turn out to be dangerous and participants in a clinical trial must receive treatment as a consequence. The Working Party views identification as "something that *must* happen under certain circumstances."[116] After all, the system of key-coded data turns on the ability to re-identify if necessary to protect a specific patient. It therefore concludes, "In this case, one can conclude that such key-coded data constitutes information relating to identifiable natural persons for all parties that might be involved in the possible identification and should be subject to the rules of data protection legislation."[117]

This analysis sweeps too broadly. Consider a scenario where a data controller maintains encrypted keys along with strong institutional safeguards to prevent access to the key-coded data unless carefully defined events occur. In that case, the party who has access to the data, but not the keys, handles information that is functionally nonpersonal information for that party.[118] In certain circumstances, therefore, the possibility of identification may be highly remote for the party who has access only to key-coded data.

A final example demonstrates how the European Union's concept of personal data skimps on analysis of whether data is reasonably likely to be identified. This illustration is drawn from Christopher Kuner's treatise on European data protection law.[119] Kuner found that the European Union's concept of "identifiability" includes a set of data that can be matched to a particular person by some party, somewhere, regardless of whether the data processor can do so.[120] His example concerns "all males over 50 living in city X who are physicians, have two daughters, listen to Verdi operas and have vacation houses in the south of France."[121] Such information is personal data "even if the data controller could not, with reasonable effort, create a link to an identifiable individual, as long as *any* party could do so."[122]

Kuner's discussion accurately reflects the current EU "state of play" on this question while also pointing to something analytically troubling about EU information privacy law. Assume that the data processor in question cannot, as

114. *Id.* at 19.
115. *Id.*
116. *Id.* at 19–20.
117. *Id.*
118. Beyond key-coded data, health care research has developed new approaches to de-identification that permit research use, including usage across multiple health care institutions. *See* Bradley A Malin et al., *Biomedical Data Privacy: Problems, Perspectives, and Recent Advances*, 20 J. AM. MED. INFORMATICS ASS'N 1, 2, 4 (2013).
119. KUNER, *supra* note 42.
120. *Id.* at 92.
121. *Id.*
122. *Id.* (emphasis in original).

Kuner posits, create a link to an individual "with reasonable effort."[123] As a further condition, assume that the data controller also institutes strong measures to keep this data secure and promises never to share it with any parties who can link it to an individual. It seems unreasonable under these conditions to require that this information receive the full set of privacy protections afforded to identified data.

Finally, this approach ultimately goes against the underlying touchstone of EU privacy law regarding identifiability. Recall that the key test in both the Directive and Proposed Regulation is whether a person is capable of identification, based on analysis of all means likely to be used. The Working Party's 2007 Opinion and Kuner's example concerning the Verdi-loving physicians demonstrate how far EU law can depart from this underlying touchstone.

2. Evaluating the U.S. Approach

There are also considerable flaws in the U.S. approach to personal information. Recall that there is no single U.S. definition of this term, but instead three approaches: the tautological, the nonpublic, and the specific-types approaches. There are two general flaws in having three available classifications. First, the presence of three definitions increases the regulatory maze and associated compliance costs for regulated entities. Second, and as a consequence of the multiple possibilities flowing from the three classifications of personal data, the same information may or may not be personal data under different statutes and in different processing contexts. Information that does not fall on a statutory list within the specific-types classification might still qualify as personal data under the nonpublic approach.

In addition, each of the three approaches in the United States has its own flaws. The tautological approach fails to define personal information or explain how to single it out. At its core, this approach simply states that personal information is personal information. As a result, this definition is unhelpful in distinguishing personal data from nonpersonal data. The process of line-drawing is likely to be based on ad hoc intuitions of regulators and judges.[124]

The initial problem with the nonpublic 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 whether it can or cannot identify a person. 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 nonpublic data that they do not want matched to this supposed public

123. *Id*.

124. This ad hoc line-drawing results in part from the approach's use of a standard rather than a rule. For further discussion of this distinction, see *supra* note 65.

information. Yet, an approach that only protects nonpublic information as PII might not preclude such combinations.

The second problem with the nonpublic approach is that it confusingly suggests that if information is public somewhere, the parties who process it are not handling regulated data. This confusion arises under the Gramm-Leach-Bliley Act.[125] As we have seen, this statute regulates "nonpublic personal information."[126] This term is defined as "personally identifiable financial information" that a consumer supplies or that is obtained in connection with a transaction involving a financial product or service.[127] In turn, "publicly available information" is defined as "any information that a financial institution has a reasonable basis to believe is lawfully made available to the general public."[128] Sources of this information can include federal, state, or local government records, or widely distributed media.[129]

This approach has the potential to mislead regarding when organizations do and do not have personal information that falls under the statutory scheme. The concept of "nonpublic personal information" may mistakenly suggest to entities that certain information does not fall under the Gramm-Leach-Bliley Act, or other federal privacy regulations, because it is available from another entity. Flaws in the regulation of data brokers and inadequate enforcement mechanisms in the applicable statutes have made a host of financial data widely accessible. For example, *The New York Times* has reported on individuals using online financial information, including credit scores, as part of the assessment of potential dates and romantic partners.[130] Yet, these data should not be considered as "nonpublic personal information" under information privacy statutes, such as the Gramm-Leach-Bliley Act. The information, while perhaps widely available to anyone with access to the Internet, a search engine, and a credit card, should not be seen as available "lawfully," which is a requirement in the FTC's Regulations.[131]

Potentially adding confusion, the law may impose obligations on organizations even for *public* information. As an example, the FTC Safeguards Rule notes that *all* customer information must be properly safeguarded once it is in the possession of an entity that falls under the Gramm-Leach-Bliley Act.[132] The Rule requires that these entities protect "the security and

125. Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999) (codified as amended in scattered sections of 12 and 15 U.S.C.).

126. 15 U.S.C. § 6801(a) (2012).

127. 15 U.S.C. § 6809(4)(A) (2012).

128. FED. TRADE COMM'N, THE GRAMM-LEACH-BLILEY ACT: PRIVACY OF CONSUMER FINANCIAL INFORMATION (2001).

129. *Id*.

130. Jessica Silver-Greenberg, *Perfect 10? Never Mind That. Ask Her for Her Credit Score*, N.Y. TIMES, Dec. 25, 2012, http://www.nytimes.com/2012/12/26/business/even-cupid-wants-to-know-your-credit-score.html.

131. *See* 16 C.F.R. § 313.3(p) (2014).

132. 16 C.F.R. § 314.1 (2014).

confidentiality of customer records and information . . . which could result in substantial harm or inconvenience to any customer."[133] Yet, the concept of "nonpublic personal information" may mistakenly suggest to regulated entities that they face far more limited obligations.

As for the specific-types approach, an initial problem is that these laws can be quite restrictive in their definition of personal information. For example, the Massachusetts data breach statute defines personal information 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.[134] This list is under-inclusive, since there are numerous other kinds of information that, independently or with a person's name, 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. Finally, and as noted above, most data breach statutes have a fixed requirement of a last name and a first name, or the initial of the first name.[135] A leak of many other types of information can reasonably be expected to cause identification of a specific individual. This variant of the specific-types approach is too rigid to adequately protect personal privacy.

COPPA, a second example of the specific-types approach, has an advantage that data breach notification laws generally lack. COPPA explicitly references FTC rulemaking as a way to expand and adapt its definition of personal information.[136] The FTC has indeed acted to expand the definition of personal information in the statute; its revised COPPA rule in 2013 further developed the statutory concept of personal information through an expansive definition of "a persistent identifier," such as a cookie.[137] This rule, as we will explain below, represents an outlier to the typical U.S. approach to personal data because of its expansive approach.

As this analysis shows, the U.S. approach suffers from numerous weaknesses. Overall, it creates inconsistencies and can leave too much information unprotected. For example, a spokesperson for the online advertising industry has stated that its "tracking doesn't violate anyone's

133. 15 U.S.C. § 6801(b) (2012).

134. MASS. GEN. LAWS ANN. ch. 93H, § 1(a) (West 2007). The other important aspect of Massachusetts data security law, beyond its data breach aspect, is that it requires affirmative steps to protect the security and confidentiality of the personal information of Massachusetts residents. 201 MASS. CODE REGS. 17.03 (LexisNexis 2013). It continues to use a static definition of PII in this aspect of its law. *Id*. at 17.02.

135. *See* SOLOVE & SCHWARTZ, *supra* note 74, at 176–78.

136. *See* 15 U.S.C. § 6501(8)(F) (2012).

137. Children's Online Privacy Protection Rule, 78 Fed. Reg. 3972, 4009 (Jan. 17, 2013) (to be codified at 16 C.F.R. § 312.2).

privacy because the data sold doesn't identify people by name."[138] Since the United States lacks both a single model and a shared understanding of "personal information," the evaluation of this claim becomes no simple matter.

The U.S. approach is also likely to lead to gaps in protection. For example, whether information can be re-identified depends upon technology and corporate practices that permit the linking of de-identified data with already identified data. 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. The U.S. definitional approach and, in particular, the specific-types approach, do not seem well equipped to function in this world of readily available data and context-specific analysis.

3. The Disjunction Between the U.S. and EU Definitions of PII

The disjunction between U.S. and EU definitions of PII raises problems regarding the harmonization of the privacy law of these two systems. To understand these difficulties, we should consider the complex legal structure for judging the permissibility of these transfers under EU law. This analysis requires examination, first of the current approach to data transfers under the Directive, and then of the suggested future approach under the Proposed Regulation. The issue of international data transfers is highly significant because the European Union is the most important bilateral trade area for the United States.[139] Indeed, the economic relationship between the United States and the European Union is the largest in the world.[140] According to one estimate from the European Commission, over half of the EU-U.S. cross-border trade in services depends on the Internet.[141] As a consequence, barriers to "information communication services" will affect not only that sector itself, but other business sectors involved in bilateral EU-U.S. cross-border trade.[142]

Under Article 25, the Directive permits transfers to "third countr[ies]," that is, countries outside of the European Union, only if these nations have "an adequate level of protection."[143] The European Union does not generally consider the United States to provide "adequate" privacy protection.[144] As a

138. *See* Julia Angwin & Tom McGinty, *Sites Feed Personal Details to New Tracking Industry*, WALL ST. J., July 30, 2010, http://online.wsj.com/news/articles/SB10001424052748703977004575393173432219064.

139. *See* Press Release, European Commission, EU-US Trade Talks: EU and US Announce 4th Round of TTIP Negotiations in March; Stocktaking Meeting in Washington D.C. to Precede Next Set of Talks (Jan. 28, 2014) ("The EU and the US make up 40% of global economic output and their bilateral economic relationship is already the world's largest.").

140. *See generally* WILLIAM H. COOPER, CONG. RESEARCH SERV., RL30608, EU-U.S. ECONOMIC TIES: FRAMEWORK, SCOPE AND MAGNITUDE (2013).

141. *See Commission Staff Working Document: Impact Assessment Report on the Future of EU-US Trade Relations*, at 8 n.11, COM (2013) 136 final (Mar. 12, 2013).

142. *Id.*

143. Data Protection Directive, *supra* note 15, at art. 25(1).

144. *See* Schwartz, *supra* note 45, at 1979-80.

consequence, negotiations took place starting in the late 1990s among different EU-U.S. institutions and private organizations to develop mechanisms for U.S. companies to meet the "adequacy" requirement of Article 25 of Directive.[145] The ensuing negotiations among a largely ad hoc "harmonization network" led to a considerable trans-Atlantic policy accomplishment.[146] The network developed a series of second-order processes by which U.S. companies can demonstrate the provision of adequate information privacy process. The policy instruments in question are the Safe Harbor, standard contractual clauses, and Binding Corporate Rules (BCR).[147] It is worthwhile to examine these second-order processes and their likely future under the Proposed Regulation before we consider how differing definitions of PII can destabilize this result.

The Safe Harbor, which was negotiated between the European Union and U.S. Department of State, creates a voluntary self-certification program for U.S. firms.[148] Its mixture of substantive standards combines EU and U.S. privacy requirements, but ends somewhat closer to the EU version of these norms.[149] The European Union has also approved two sets of standard contractual clauses, which simplify the process of crafting data transfer agreements by providing "off-the-rack" terms for agreements.[150] The development of standard contractual clauses involved a significant role for a non-governmental organization, the International Chamber of Congress, located in Paris.[151] BCRs provide another way to meet the Directive's adequacy requirement.[152] Through BCRs, an organization pledges to meet the Directive's standards in its data processing and promises its cooperation with EU data protection authorities.[153]

---

145. *See* NEWMAN, *supra* note 20, at 74–82.

146. *See* Schwartz, *supra* note 45, at 1991 (citing ANNE-MARIE SLAUGHTER, A NEW WORLD ORDER 5, 15, 20, 59, 162 (2004) (introducing the concept of "harmonization networks" that develop when different nations' regulators work together to harmonize and adjust domestic law to achieve efficiency and mutually acceptable outcomes)).

147. *See id.* at 1980–84.

148. *See* Issuance of Safe Harbor Principles and Transmission to European Commission, 65 Fed. Reg. 45,666 (July 24, 2000).

149. *See* Schwartz, *supra* note 45, at 1981.

150. Commission Decision (EC) 2001/497 of 15 June 2001 on Standard Contractual Clauses for the Transfer of Personal Data to Third Countries, under Directive 95/46/EC, 2001 O.J. (L 181) 19, 19–20; Commission Decision (EC) 2004/915 of 27 Dec. 2004 Amending Decision 2001/497/EC as Regards the Introduction of an Alternative Set of Standard Contractual Clauses for the Transfer of Personal Data to Third Countries, 2004 O.J. (L 385) 74, 74–75.

151. For an analysis of the International Chamber of Commerce (ICC) Model Clauses, see Lingjie Kong, *Data Protection and Transborder Data Flow in the European and Global Context*, 21 EUR. J. INT'L L. 441, 449 (2010).

152. *See* Article 29 Data Protection Working Party, Working Document: Transfers of Personal Data to Third Countries: Applying Article 26(2) of the EU Data Protection Directive to Binding Corporate Rules for International Data Transfers, 11639/02/EN/WP 74, at 5–6 (June 3, 2003).

153. For a list of companies that have been approved under the BCRs, see *List of Companies for Which the EU BCR Cooperation Procedure Is Closed*, EUR. COMM'N, http://ec.europa.eu/justice/data-protection/document/international-transfers/binding-corporate-rules/bcr_cooperation/index_en.htm (last visited Mar. 23, 2014).

The Proposed Regulation generally continues the approach of the Directive to data transfers outside of the European Union. Like the Directive, it allows these transfers when the personal data will receive "an adequate level of protection."[154] The language of "adequacy" should not lead one to imagine a low standard of protection. Under the Proposed Regulation, as well as the Directive, the goal is to protect the fundamental right of data protection on a global basis.

The Proposed Regulation is also more flexible than the Directive in certain regards. The Proposed Regulation notes that the Commission makes an adequacy determination through examination of "the third country, or a territory, or a processing sector within that third country, or the international organisation in question."[155] This language suggests that smaller geographical areas, such as individual states or particular companies, might receive an adequacy determination from the Commission. As a further indication of flexibility, the Proposed Regulation permits transfers if there is a use of "appropriate safeguards . . . in a legally binding instrument" as part of the transfer, or the use of one of eight possible exceptions.[156] Finally, the Proposed Regulation grandfathers in acceptance of the Safe Harbor, standard contractual clauses, and BCRs.[157]

Under either the Directive or the Proposed Regulation, however, the resulting analysis is complicated by the differences in definitions of PII in the European Union and United States. Just as the differing legal definitions of PII within the United States raise compliance costs for U.S. companies, the differing approaches between the European Union and United States further heighten regulatory uncertainty. Information that is "identifiable" in the European Union (hence subject to Article 25 of the Directive) or "indirectly identified" (hence subject to the Proposed Regulation) may not be subject to privacy law in the United States. As a consequence, international companies face complex legal decisions when designing software, services, or devices for use in both the European Union and United States.

Furthermore, the different definitions of PII threaten the status quo around second-order mechanisms for allowing data transfers. These are the Safe Harbor, standard contractual clauses, and BCRs. If the European Union and United States cannot agree on a definition of PII, the most basic unit of information privacy law, these processes must be seen as essentially instable. Each is based on the agreement around procedures protecting personal information.

As an illustration of the destabilization of the existing EU-U.S. privacy status quo, consider a 2013 speech by Viviane Reding, the Vice-President of

154. *Proposed Regulation*, *supra* note 18, at art. 41(1).
155. *Id.* at art. 41.
156. *Id.* at art. 42(1).
157. *Id.* at art. 42(2)–(3).

the European Commission.[158] In her address, Reding first pointed to the importance of data protection reform for the 500 million citizens of the European Union.[159] As part of the necessary reform measures, Reding identified the "need to ensure that the same rules apply to all businesses providing services to EU residents."[160] She explicitly called for "[n]on-European companies, when offering services to European consumers," to "apply the same rules and adhere to the same levels of protection of personal data."[161] Reding then turned to the question of "new technologies which allow data to be made anonymous or to be processed based on an identifier—a pseudonym—rather than the person's name."[162] Here, she welcomed the use of pseudonyms rather than actual names as in the interest of citizens, but then cautioned: "Pseudonymous data is personal data. It relates to an identified or identifiable natural person and has to be protected."[163] Reding also warned that the European Union should be vigilant in preventing companies attempting to use the category of "pseudonymous data" as "a Trojan horse at the heart of the [Proposed] Regulation, allowing the non-application of its provisions."[164]

Reding's analysis is truncated. The categories of pseudonymous and anonymous data are complex, and data that are not processed under someone's name but pursuant to another associated identifier do not simply qualify as "identified or identifiable." This speech can be seen as a harbinger of battles to come around the definition of PII. Reding's discourse also creates a strict benchmark for future scrutiny of any second-order harmonization mechanism; she suggests that these policy instruments must meet "the same rules" and "same levels of protection of personal data" as EU data protection law.[165]

There is ample additional proof, beyond this speech, of an emerging destabilization of the EU-U.S. privacy status quo. Much of this discontent is directed toward the Safe Harbor, the mechanism that allows organizations that pledge to meet EU data protection requirements to self-certify.[166] Jan-Phillip Albrecht, the member of the European Parliament who oversees the Proposed

158. Press Release, Viviane Reding, Vice-President of the European Commission, EU Data Protection Rules: Better for Business, Better for Citizens (Mar. 26, 2013), *available at* http://europa.eu/rapid/press-release_SPEECH-13-269_en.htm.

159. *Id.* at 3.

160. *Id.* at 5.

161. *Id.*

162. *Id.* at 3.

163. *Id.*

164. *Id.*

165. *Id.* at 5.

166. *See* Damon Greer, *Safe Harbor May Be Controversial in the European Union, But It Is Still the Law*, PRIVACY ADVISOR (Aug. 27, 2013), https://www.privacyassociation.org/publications/safe_harbor_may_be_controversial_in_the_european_union_but_it_is_still_the (last visited Mar. 23, 2014) (describing EU discontent with Safe Harbor mechanism); Lothar Determann, *Data Privacy in the Cloud: A Dozen Myths and Facts*, COMPUTER & INTERNET LAW., Nov. 2011, at 1, 4 ("The US-EU safe harbor program has been heavily criticized over the years, and the head of a data protection authority in one German state has even called for a rescission of the program.").

Regulation in the Parliament, has formally recommended that the European Union discontinue this mechanism two years after approval of the Regulation.[167] Subsequent to Albrecht's recommendation, another factor increased the dissatisfaction with the Safe Harbor: the 2013 revelations by Edward Snowden of the U.S. National Security Agency's surveillance of international communications.

Following the Snowden revelations and a call by German data protection commissioners for suspension of the Safe Harbor, Commission Vice President Reding announced a plan to carry out a full review of this EU-U.S. agreement.[168] The review's initial outcome was released in November 2013 and argued that "the current implementation of the Safe Harbour cannot be maintained."[169] It called for a series of improvements to address both "structural shortcomings related to transparency and enforcement, the substantive Safe Harbour principles and the operation of the national security exception."[170] Adding its voice to the discussion, the European Parliament passed a non-binding resolution in March 2014 calling for the suspension of the Safe Harbor.[171]

In this uncertain environment, the inconsistent definitions of PII pose a significant additional threat to continuing acceptance of the second-order harmonization instruments. Privacy law in the United States and the European Union have vastly different jurisdictional scopes, and these differences persist despite mechanisms, such as the Safe Harbor agreement, that smooth differences in U.S. and EU privacy law.

III.
PII 2.0

The existing definitions of personal information, whether in the European Union or United States, are problematic. Nonetheless, information privacy law should not abandon the concept of PII. If it took this step, privacy law would be left without a means for establishing coherent boundaries on necessary

167. *Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)*, at 58–59 (Dec. 17, 2012), *available at* http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/pr/922/922387/922387en.pdf.

168. Press Release, Informal Justice Council in Vilnius, MEMO/13/710 (July 19, 2013), *available at* http://europa.eu/rapid/press-release_MEMO-13-710_en.htm.

169. *Communication from the Commission to the European Parliament and the Council, Rebuilding Trust in the EU-US Data Flows*, at 6, COM (2013) 846 final (Nov. 27, 2013), *available at* http://ec.europa.eu/justice/data-protection/files/com_2013_846_en.pdf.

170. *Id.*

171. European Parliament, *US NSA Surveillance Programme, Surveillance Bodies in Various Member States and Impact on EU Citizens' Fundamental Rights and on Transatlantic Cooperation in Justice and Home Affairs*, Recommendations 40–41 (Mar. 12, 2014), http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P7-TA-2014-0230.

regulation. Therefore, we have reconceptualized the current standards in the European Union and United States through our model of PII 2.0. In this Part, we present this concept, which we introduced in a previous article, and further develop the idea by showing how it can bridge significant differences in EU and U.S. privacy law.

A. An Explanation of PII 2.0

PII 2.0 places information on a continuum. On one end of the continuum, there exists no risk of identification. At the other end, an individual is fully identified. We divide this continuum 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. Because these categories do not have hard boundaries, we define them in terms of standards—that is, as open-ended benchmarks rather than hard-edged rules.

1. Identified Data

Information refers to an *identified* person when it singles out a specific individual from others. Put differently, ascertaining a person's identity makes her identified. 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 Government Accountability Office, Office of Management and Budget, and the National Institute of Standards and Technology associate this concept with information that distinguishes or traces a specific individual's identity.[172] In the European Union, the Working Party states that a person is identified "when, within a group of persons, he or she is 'distinguished' from all other members of the group."[173] This definition also follows that of member states, which, as we have seen, assess whether information relating to a person has determined her specific identity.[174] To return to an example from Rosemary Jay about UK data protection law, a person is identified "where there is sufficient information either to contact him or to recognise him by picking him out in some way from others and know who he/she is."[175]

EU data protection law also contains special protections for sensitive data, and our model of PII 2.0 would leave this special designation in place for the European Union and in such sectors in the United States that recognize it. The Directive specifies that "sensitive data" are data about "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union

172. ERIKA MCCALLISTER ET AL., NAT'L INST. OF STANDARDS & 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 (2008).

173. Article 29 Data Protection Working Party, *supra* note 90, at 12.

174. *See supra* notes 21–23 and accompanying text.

175. JAY, *supra* note 22, at 172.

membership, . . . health or sex life."[176] In Europe, these data receive stronger protections a priori than other types of data.[177] The Proposed Regulation also recognizes a similar category. As its Recital 41 states, "Personal data which are, by their nature, particularly sensitive and vulnerable in relation to fundamental rights or privacy, deserve specific protection."[178] Article 9 of the Proposed Regulation provides protections for such "special categories of personal data."[179]

Such a category does not exist as a general matter in U.S. privacy law. Yet, U.S. law does extend heightened protection to certain data through specific laws and regulations, such as the Health Insurance Portability and Accountability Act (HIPAA). But for the most part, U.S. law does not globally recognize types of data that receive heightened protection across various laws akin to EU-style "sensitive data."

PII 2.0 does not solve the divergence in conceptual approaches toward protecting special categories of personal information in the European Union and United States. There is only so much harmonization possible through PII 2.0 alone, and the different approaches to sensitive information are deeply embedded in the EU and U.S. legal systems. Thus, under PII 2.0, identified data that fall into the sensitive category would continue to receive special protections pursuant to EU law and where mandated by U.S. sectoral privacy law.

There are also 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 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. Thus, within the "identified" information section of the PII 2.0 continuum, we propose a new subcategory for this type of identifiable information.

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 time during which information is to be stored, the likelihood of future relevant technology development, and parties' incentives to link identifiable data to a specific person. It should also consider steps that a company takes to keep information from being linked to any specific individual.

176. Data Protection Directive, *supra* note 15, at art. 8.
177. *See id.*; *Proposed Regulation*, *supra* note 18, at pmbl. ¶ 41.
178. *Proposed Regulation*, *supra* note 18, at pmbl. ¶ 41.
179. *Id.* at art. 9.

2. Identifiable Data

Information in the middle of the PII 2.0 risk continuum relates to an *identifiable* individual when specific identification, while possible, is not a significantly probable event, but there is nonetheless some non-remote possibility of future identification. The risk level for such information is low to moderate. Information of this sort should be regulated differently from that important subcategory of nominally identifiable information, mentioned above, in which linkage to a specific person has not yet been made, but where such a connection is likely. Such nominally identifiable data should be treated the same as identified data. This is a more targeted approach than European Commission Vice President Reding's choice to treat all pseudonymous data as identified information.[180]

An example of *identifiable* information under the PII 2.0 model would be the key-coded medical data that the Working Party discussed in its "Opinion on the Concept of Personal Data."[181] Some or all of this information might never be identified. Depending on the risk scenario, there may be only a remote chance of future linkage to a specific person. As a further example, Kuner's discussion of the Verdi-loving physician may represent merely identifiable information under PII 2.0.[182] Kuner's hypothetical leaves much open regarding the "data controller." We know only that the data controller himself cannot identify the person to whom the information relates. If the data controller also has strong measures in place to protect the data from exposure to others, the PII 2.0 model would classify the information as identifiable, but not identified.[183]

For an example from the United States regarding "identifiable" but not "identified" information, we turn to the FTC Staff Report, *Protecting Consumer Privacy in an Era of Rapid Change*.[184] This Report considers the issue of when information is "reasonably linkable" to a person.[185] Citing to our previous work on PII 2.0, the FTC noted that businesses can sometimes re-identify non-PII data and often have incentives to do so.[186] It argued that if companies take three specific steps to minimize linkability, the information should be viewed as non-PII.[187] First, the company must use reasonable means to ensure that the data is de-identified, or cannot be tied to a specific

180. *See* Reding, *supra* note 158, at 3.

181. *See* Article 29 Data Protection Working Party, *supra* note 90, at 19.

182. *See* KUNER, *supra* note 42, at 92.

183. An example is health care data that is subject to de-identification and shared for health care research. Regarding the ongoing debate to ensure robustness of de-identification models, see Deven McGraw, *Building Public Trust in Uses of Health Insurance Portability and Accountability Act De-identified Data*, 20 J. AM. MED. INFORMATICS ASS'N 29, 32 (2013).

184. FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE (2012).

185. *See id.* at iv.

186. *Id.* at 20.

187. *Id.* at 20–21.

consumer.[188] Second, the company must publicly promise that it will use the data only in de-identified fashion and not attempt to re-identify it.[189] Finally, if the company makes the de-identified data available to other companies, it must contractually bind the other entities to avoid re-identifying the data and to engage in reasonable oversight to monitor compliance with these contracts.[190] These steps demonstrate a practical policy for maintaining information in the identifiable category.

3. Non-identifiable Data

At the other end of the risk continuum, *non-identifiable* information carries only a remote risk of identification. Such data are not 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.

A simple example of non-identifiable information is high-level information about the populations of the United States, China, and Japan, and their relative access to telecommunications. At an abstract level, this information refers to persons; it is not merely data about the physical or manmade world, such as the sky is blue or that Route 95 goes through New Haven, Connecticut. Yet, this information also cannot be linked to a specific person.

Practical methodologies now 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.[191] 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.[192]

In addition, computer scientists' ongoing work in developing secure software offers useful lessons for evaluating the risk of re-identification. The relevant focus includes (1) the nature of internal and external threats to a de-identified data asset, and (2) the effectiveness of possible countermeasures to

188. *Id.* at 21.

189. *Id.*

190. *Id.*

191. *See* KHALED EL EMAM, GUIDE TO THE DE-IDENTIFICATION OF PERSONAL HEALTH INFORMATION 151–58 (2013); Khaled El Emam, *Heuristics for De-Identifying Data*, IEEE SECURITY & PRIVACY, July/Aug. 2008, at 58; Khaled El Emam, *Risk-Based De-Identification of Health Data*, IEEE SECURITY & PRIVACY, May/June 2010, at 64 [hereinafter El Emam, *Risk-Based De-Identification*].

192. El Emam, *Risk-Based De-Identification*, *supra* note 191, at 66. For a further elaboration, see EL EMAM, *supra* note 191.

those threats.[193] Data security has its own developed methodologies for assessing risks to software and computer systems.[194]

B. PII 2.0 and Fair Information Practices (FIPs)

Our reconceptualized notion of personal information places greatest emphasis on the risk level associated with potential identification. PII 2.0 conceives of identifiability as a continuum of risk rather than as a simple dichotomy. Unlike the EU's simple "on" or "off" switch for information privacy law, our model envisions a more modulated approach. 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) creation of processing systems that the concerned individual can know about and understand (transparent processing systems); and (7) security for personal data.[195] When information refers to an *identified* person, all of the FIPs generally should apply.

As for the category of *identifiable* information, it is not appropriate to treat such information as fully equivalent to identified information. 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.

In thinking about FIPs for identifiable data, the easiest starting point is to eliminate inapplicable categories. 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 perhaps decrease rather than increase privacy by requiring that all such data be associated with a specific person. This result follows because entities would need to maintain an ongoing connection between the individual and the identifiable information to allow that individual to exercise her rights of notice, access, and correction. In this fashion, the law's implementation could force the transformation of *identifiable* data into *identified* data. Article 10 of the Proposed Regulation explicitly seeks to avoid

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

194. *See id.*

195. *See, e.g.*, Org. for Econ. Co-operation & Dev., *Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, OECD Doc. C(80)(58)/FINAL, as amended on 11 July 2013 by C(2013)79, *available at* http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf.

this result. It provides that a data controller is not obligated to collect further information in order to identify the data subject for the mere purpose of complying with the Proposed Regulation.[196]

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 would cripple socially productive uses of analytics that do not raise significant risks of individual privacy harms.[197]

Some of these uses of analytics are consumer-oriented and some are not, but the benefit to the public is often clear. As an example of a socially productive and consumer-focused service, Google Flu Trends provides geographic information on the spread of the influenza virus based upon search queries entered into Google's search engine.[198] While Flu Trends is a prominent public example of analytics, it represents a modest start on discovering correlations through a data-driven approach.[199]

Among non-consumer use of analytics, analysis of large data sets plays an increasingly important role in health care research, the management of physician performance and clinical metrics, data security, and fraud prevention.[200] Additionally, the realm of health care research has shifted away from traditional clinical trials that follow specific patients toward informational research that analyzes large data and biological-sample sets. The Institute of Medicine explains these new "information based" forms of inquiry as "the analysis of data and biological samples that were initially collected for diagnostic, treatment, or billing purposes, or that were collected as part of other research projects."[201] This technique, centered on analytics, is widely used today in categories of research including epidemiology, health care, and public health services. These information-based forms of health research "have led to significant discoveries, the development of new therapies, and a remarkable improvement in health care and public health."[202]

---

196. *Proposed Regulation*, *supra* note 18, at art. 10.

197. 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, 00062/10/EN/WP 173, at 3, (July 13, 2010). The question remains as to how successful this concept can be in a system that treats identified and identifiable data as equivalents. *See* Article 29 Data Protection Working Party, Opinion 4/2007 on the Concept of Personal Data, 01248/07/EN/WP 136 (June 20, 2007).

198. Jeremy Ginsberg et al., *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012, 1014 (2009).

199. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 106, at 52–53.

200. *Id*. at 59–61.

201. INST. OF MED. OF THE NAT'L ACADS., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 112 (Sharyl J. Nass et al. eds., 2009).

202. *Id.* at 113. For an illustrative case study, see David C Kaelber et al., *Patient Characteristics Associated with Venous Thromboembolic Events: A Cohort Study Using Pooled Electronic Health Record Data*, 19 J. AM. MED. INFORMATICS ASS'N 965 (2012).

As noted above, while all FIPs should not apply to identifiable data, there are three FIPs that are applicable in this context: 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."[203] Identifiable information 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. There are social costs to both under-protecting and over-protecting personal information.

The transparency FIP calls for the creation of data processing systems that are open and understandable to affected individuals. There are a number of reasons that this FIP is important. 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."[204] More recently, the American Civil Liberties of Northern California trenchantly noted, "Transparency requirements . . . incentivize companies to better protect consumer privacy."[205] Second, identifiable information can provide great value to companies and consumers who furnish the raw information for the new age of Big Data. Transparency about the collection of identifiable information will heighten awareness about data flows among all parties, both consumers and corporations. It will thereby improve the position of consumers who have preferences about the collection and further use of data—even should that information merely be identifiable.

Finally, data quality is an 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, the greater the potential harm to individuals, the more precise the data and its processing must be. Some decisions matter more than others, and the stakes are low when the issue is whether or not one receives a coupon for a dollar discount on a case of mineral water. More accuracy is required for a data system that processes information to decide whether or not one receives a mortgage and calculates 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 is potentially greater, the data

203. SOTTO, *supra* note 79, at § 14.01.

204. LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 92 (1914).

205. ACLU OF CALIFORNIA, LOSING THE SPOTLIGHT: A STUDY OF CALIFORNIA'S SHINE THE LIGHT LAW 4 (2013), *available at* https://www.aclunc.org/publications/losing-spotlight-study-californias-shine-light-law.

quality requirements must be higher. In the context of *identifiable* information, data quality also requires good practices of information handling. In particular, it requires that companies pay attention to the use and processing 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 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.[206] An example is information used in health care research. Access to such data should be accompanied by legal 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.

C. PII 2.0 and EU Privacy Law

Our model of PII 2.0 has elements that are distinct from EU law and U.S. law alike. The Working Party would treat all information collected by a data controller as identifiable, and hence subject to full protection, so long as the ultimate intention is to link some of these data to individuals. Indeed, none of the information collected may ever be identified; this result is demonstrated by the hypothetical example of the Working Party concerning the "key-coded data." Further, the Kuner example of the Verdi-loving physician shows how EU law has moved away from a requirement that the party who can link data to a specific individual be reasonably likely to obtain the information. In contrast, and as noted earlier, the Proposed Regulation's Recital 24 adopts a context-specific analysis and states, for example, that "location data . . . need not be considered as personal data in all circumstances."[207]

PII 2.0's distinction from U.S. law is also clear. U.S. law tends only to protect identified information. State data breach notification law generally takes the approach that unless a last name and first name or first initial are disclosed, an organization is not required to inform an affected individual about leaked information. The tautological approach is of even less help. It states that "personally identifiable information" is "information that identifies a person." This language seems to suggest that the identification must have already taken place and does not depend on events reasonably likely to occur.

At the same time, PII 2.0 attempts to align U.S. and EU privacy law by using concepts derived from each. From the U.S. approach, PII 2.0 takes a more harm-based approach. Like U.S. law, it gives data about identified individuals the most protection. Like EU law, PII 2.0 recognizes that

206. Paul Ohm, *Broken Promises of Privacy*, 57 UCLA L. REV. 1701, 1741–42 (2010).
207. *Proposed Regulation*, *supra* note 18, at pmbl. ¶ 24.

identifiable data still deserves protection and should be included within the definition of PII.

But is PII 2.0 truly compatible with the EU approach? Upon initial reflection, one might expect the answer to be "no." PII 2.0 provides only some of the FIPs to certain kinds of data that EU privacy law would protect with all of the FIPs. The EU approach also applies its FIPs uniformly to all PII, while PII 2.0 permits variations in protection. Thus, on the surface, PII 2.0 might appear to weaken EU privacy protection and contravene its goal of providing a uniform and high level of privacy protection to data to respect individuals' fundamental right to privacy.

In our view, however, PII 2.0 is not only fully compatible with the EU approach, it is consistent with its underlying philosophy and effectively furthers its goals. As a larger point, the concept of sensitive data shows how the European Union already supports different categories of data with different levels of protection. The Directive identifies a special category of data called "sensitive data" and provides it with stronger protections than other types of data.[208] The Proposed Regulation contains a similar category.[209] Thus, the EU approach already diverges from uniformity when different levels of protection will better protect individuals' right to privacy, and would in this way align with the PII 2.0 proposal.

In addition, the Proposed Regulation and the Working Party indicate that the full requirements of EU data protection law need not apply to all types of personal data, be it identified or identifiable information. At present, however, while this evidence does not represent the conventional wisdom, it provides some support for the evolution of the majority view. As mentioned previously, the Proposed Regulation recognizes that applying its full requirements to identifiable data would create, at least at times, the perverse result of obligating organizations to collect more personal data in order to authenticate the data subjects. The drafters therefore wisely included Article 10, which provides that data controllers need not collect more personal information to identify the data subject for the mere purpose of complying with the Proposed Regulation.[210]

The logic of Article 10 is impeccable—it recognizes that identifiable information should not and cannot be regulated in the same manner as identified information. Thus, while the Proposed Regulation does not specifically create two classes of personal data with differing requirements, Article 10 would permit such results. Yet, Article 10 is no panacea. It is vague regarding (1) the types of personal data to which it would apply and (2) the provisions of the regulations with which a data controller need not comply if it had such information. Thus, while Article 10 recognizes that identified personal

208. Data Protection Directive, *supra* note 15, at art. 8.
209. *Proposed Regulation*, *supra* note 18, at art. 9.
210. *Id.* at art. 10.

information should be regulated differently from identifiable information, it is an incomplete solution. PII 2.0 addresses this need for nuance regarding classification of personal information.

The Working Party's opinions helped develop the EU's expansionist views of personal data. We have noted some of the shortcomings in this approach, such as redefining "identified" as decision making based on a person's specific characteristics or whether *some* parties in a database might be identified. Yet, in its 2011 opinion on geolocation data, the Working Party found some information deserved a lighter set of FIPs because it posed a lesser privacy risk.[211]

The Working Party's ultimate conclusions about Wi-Fi routers demonstrated a modest, initial step that may lead, one day, to evolution of the EU's view toward PII. Initially, the Working Party broadly stated that a data controller should treat "all data about WiFi routers as personal data."[212] Even when "in some cases the owner of the device currently cannot be identified without unreasonable effort," a Wi-Fi access point should be viewed as personal data.[213] It reached this conclusion because the information can be indirectly identified in certain cases. Thus, the opinion of the Article 29 Working Party did not demonstrate flexibility in the definition of "personal data." Its starting point was that a Wi-Fi MAC address in combination with location information constituted "personal data." Yet, it also found that this information posed a "lesser threat to the privacy of the owners of these access points than the real-time tracking of the locations of smart mobile devices."[214] Due to this "lesser threat," the Working Party took some initial steps on the path to PII 2.0. It called for a less rigorous opt-out mechanism, rather than an automatic opt-in, as well as a lighter notice requirement, and it implied that access for the affected individual need not be provided if provision would require collection of additional information to authenticate the Wi-Fi access point owner.[215]

Thus, PII 2.0 is consistent with at least some existing strands in EU information privacy law. Most importantly, PII 2.0 enhances the protection of privacy. It creates an incentive for companies to keep information in its 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. They will also be less

211. Article 29 Data Protection Working Party, Opinion 13/2011 on Geolocation Services on Smart Mobile Devices, 881/11/EN/WP 185, at 7 (May 16, 2011).

212. *Id.* at 11.

213. *Id.*

214. *Id.* at 16.

215. *Id.* at 16, 18.

likely to develop strong contracts with outside parties to keep shared information de-identified, as the FTC proposes.[216]

Beyond the incentive to keep data de-identified, PII 2.0 enhances privacy. Administering certain FIPs requires that data be identified, and keeping data in identified format can create privacy risks. Providing individuals with access to their data, for example, requires that the information be kept in identified form. But by keeping the data in identified form, there is an increased risk from a potential data security breach. If data is not kept in this form, data processors would not know to whom to provide access. The PII 2.0 approach would avoid these potential pitfalls by incentivizing de-identification.

By contrast, when a breach involves only *identifiable* data, the harm that the information can cause to individuals is much less likely to occur. Harm can only occur when the party who obtains the data also knows how to identify it. Although identification of some data may be theoretically possible, individuals with unauthorized access to it may lack the resources or knowledge to do so. Indeed, media accounts of at least one supposed triumph of re-identification proved overstated. Professor Daniel Barth-Jones debunked the popular account of a 1997 incident involving William Weld, then Governor of Massachusetts, whose medical data was purportedly easily identified through the use of voter registration rolls in Cambridge, Massachusetts.[217] Barth-Jones demonstrated that Weld's re-identification rested on certain unusual aspects of the population demographics of Cambridge, Massachusetts, including a notable scarcity of registered Republicans.[218] Barth also argued in favor of the robust nature of the protections of the Privacy Rule, which the federal government issued pursuant to HIPAA.[219]

Keeping data in de-identified form prevents harms from inappropriate access by employees or others. The risk of inappropriate access makes it harder to engage in new uses of data, which is why the FTC seeks to have companies contractually prohibit downstream recipients from re-identification of shared data. Beyond these legal requirements, of course, the mere status of information in de-identified form creates obstacles to identification by raising technological barriers and imposing costs for outsiders.

Therefore, for the goal of protecting privacy, it is far preferable to keep data in identifiable rather than identified form. PII 2.0 encourages keeping data in this format, while the EU approach to PII discourages keeping data merely identifiable. For these reasons, PII 2.0 would strengthen privacy protection in

---

216. *See* FED. TRADE COMM'N, *supra* note 184, at 21.

217. Daniel C. Barth-Jones, The "Re-identification" of Governor William Weld's Medical Information (July 24, 2012) (working paper) (*available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2076397).

218. *Id.*

219. *Id.*

the European Union and resolve some of the ambiguities of EU data protection law.

CONCLUSION

Despite the divergence between the concepts of personal data in the United States and the European Union, the differences between the two approaches can be reconciled to a certain extent. PII 2.0 rationalizes the currently inconsistent U.S. approach to defining personal data. It also is compatible with basic principles of U.S. privacy law by focusing on the risk of harm to individuals. PII 2.0 is consistent as well with the acknowledgment of EU privacy law of the need to provide different categories of information with different kinds of protection. In the European Union, it would provide for more tailored and nuanced protection. Most importantly, in both the European Union and United States, it would enhance the protection of privacy by creating an incentive for companies to keep information in the least identifiable form. Thus, PII 2.0 is the ideal starting point toward reconciling these divergent bodies of law.