Fordham Urban Law Journal

Volume 46
Number 2 *Artificial Intelligence and Predictive Algorithms: Why Big Data Can Lead To Big Problems*

Article 2

2019

Beyond the "Nature" of Data: Obstacles to Protecting Sensitive Information in the European Union and the United States

Müge Fazlioglu

BEYOND THE "NATURE" OF DATA: OBSTACLES TO PROTECTING SENSITIVE INFORMATION IN THE EUROPEAN UNION AND THE UNITED STATES

Müge Fazlioglu[*]

ABSTRACT

Privacy and data protection laws in both the European Union and the United States impose heightened obligations on data controllers and processors that handle data deemed to be of a "sensitive" nature, such as health information, financial information, and information concerning minors. But the central assumptions that underlie these special protections, imposed through various laws on both sides of the Atlantic, have lagged behind modernization and technological advancements.

The current scheme of sensitive data protection faces three primary obstacles, which, taken together, show that prioritization of sensitive information is no longer sufficient on its own to deal with the most significant privacy risks to individuals. The first challenge is the rapid growth in data-collecting technologies, which have led to the emergence of new types of data. This new information can range from behavioral data produced by online activity and collected by social networking sites, to geolocation data produced by cell phone and other smart devices that monitor users' movements and activities. The second challenge, closely related to the first, is that non-identifiable, non-sensitive types of data can still be linked to an

---

[*] Senior Westin Research Fellow, International Association of Privacy Professionals (IAPP); Ph.D., Law and Social Science, Maurer School of Law, Indiana University-Bloomington (2017). I wish to thank my dissertation advisor, Professor Fred Cate, for his invaluable feedback on earlier versions of this Article. I also appreciate the other readers on my dissertation committee, Beth Cate, Mark Janis, and Michael McGregor, who offered useful suggestions to me. At IAPP, I owe a debt of gratitude to my supervisors, Rita Heimes and Omer Tene, for their helpful comments, as well as to my colleague Mitchell Noordyke, for lending his skills at a critical time. Last, but not least, I am grateful to Kyle Heatherly for his indefatigable love and support.

identity or used to reveal or infer sensitive data using innovative analytic techniques. Many datasets that attempt to anonymize or de-identify data — that is, tools that expunge personally-identifying information such as names, addresses, or dates of birth — may still be combined with other datasets to re-identify individuals. The third and final obstacle is that the "sensitivity" of a given piece of information can change depending on the context of its use. This means that the sensitivity level of a piece of data is not solely a function of its nature or type, but also of the way in which it is used or the ends for which it is utilized. Thus, privacy and data protection laws that assume sensitivity is a static quality of certain data types are not in sync with the reality of data use.

This Article concludes by suggesting ways policymakers can rethink the prioritization of sensitive information and address newly-emerging risks to information privacy. Relying entirely on the sensitivity level of a piece of data to determine the risks associated with it will fall short of adequately protecting data subjects' privacy. Data controllers and regulators must therefore consider other factors, in addition to the category of data or its sensitivity level, such as whether it can be used in combination with other publicly-available data to uniquely identify a person, the likelihood it can be linked to or reveal sensitive data about a person, and the context of the data use, when determining the risks posed by data processing. Laws also ought to extend protection to newly-emerging types of data that are sensitive in nature, such as web-browsing histories and longitudinal location data. Law and policymakers should ultimately move beyond the category-based regulation of data to effectively protect privacy today.

TABLE OF CONTENTS

INTRODUCTION

Privacy and data protection laws in both the European Union and the United States impose conditions on the processing of certain kinds of personally identifying information, referred to as *sensitive data*, *sensitive information*, or *special categories of information*. The sensitivity of a given piece of information is often used to determine how much legal protection should be afforded to it. This is because greater consequences are likely to result from the misuse of sensitive information compared to misuse of less sensitive or non-sensitive information. Thus, information sensitivity is associated with risk. Indeed, protecting sensitive data has become one of the most important issues in the domain of risk management in recent years.

The categories of information that data protection and privacy laws have tended to recognize as sensitive include health data, financial data, and other types of information considered to be of an intimate or personal nature. The exposure of these types of information is thought to bring about the most severe kinds of privacy harms. And yet, we live in an era where almost any piece of data about a person, sensitive or not, can be linked to their identity. Further, innocuous bits of information can be aggregated from multiple sources that, when observed as a whole, reveal sensitive attributes about a person. Laws that treat certain types of information as warranting heightened legal obligations, therefore, may ultimately fail to adequately protect privacy if they ignore how non-sensitive data can be linked to sensitive data. This Article focuses on understanding why making this link is so vital and what can be done to incorporate a more nuanced conceptualization of the categorization of sensitive data in privacy and data protection law and policymaking.

Ultimately, the reality of a technologically- and data-dependent world means that lawmakers face an uphill battle and must be willing to continuously evaluate and potentially expand the list of data types that receive heightened legal protections. Additionally, lawmakers must address the way in which seemingly innocuous pieces of information can be connected to sensitive types of data, as

developments in technology and big data have diminished the utility of the longstanding divide between sensitive and non-sensitive data. In line with this concern, policymakers must also reconsider the effectiveness of current approaches to data protection, most of which prioritize information solely according to its *nature*, *type*, or *category*. Additionally, lawmakers should account for whether the data can be readily combined with other publicly-available information to uniquely identify a person, the likelihood that it can be linked to or reveal sensitive information, and the context of data use when determining what legal protections should be afforded to it.

Before analyzing these obstacles and discussing ways of addressing them, Part I of this Article explains why certain types of information, such as health or financial data, merit heightened legal protections and impose heavier obligations on data processors, collectors, and providers. This is done by focusing on the levels of risk associated with certain types of data across the United States and European Union. Part II of this piece dives into the myriad obstacles that exist in protecting sensitive data, examining how current legal protections are insufficient. Additionally, Part II discusses why proper protections for sensitive information are so vital. Part III suggests various ways in which policymakers ought to rethink sensitive data, and the kind of impact this rethinking can have on privacy rights and data protections.

I. PRIORITIZING SENSITIVE DATA

This Article begins by explaining why the severity of privacy harms that arise from misuse of sensitive data are often greater than the privacy harms associated with misuse of non-sensitive types of data. It then examines the reliance of privacy and data protection laws on the *type*, *nature*, or *category* of information to prioritize protection. Laws that prioritize sensitive information rely on assumptions about data that are being invalidated by innovative developments in technology and data use. More specifically, a purely categorical approach to privacy and data protection fails to recognize the risks generated by the expansion in data collection practices and technologies. Assumptions made by policymakers and legislators in both the European Union and the United States about managing the risks associated with certain types of data thus need to be revisited with these emerging threats to privacy in mind.

A. Why Processing Sensitive Data Is Assumed to Entail More Risk

The sensitivity of a given piece of information is often defined as a function of the magnitude and severity of the risks associated with its processing. For example, one set of guidelines on handling sensitive health information defines it as "information that carries with it unusually high risks in the event of disclosure."[1] Similarly, experimental studies surrounding privacy attitudes and behaviors have demonstrated that individuals' perceptions of information sensitivity are positively correlated with their perceptions of risk or exposure, or the heightened need for privacy.[2] It is well-recognized that the processing of sensitive information, if not handled properly, "can lead to significant forms of harm [to individuals] . . . [and] is the kind that exposes the data subject to a high probability of such harm."[3]

Indeed, a key difference between sensitive and non-sensitive information is the level of risk associated with disclosure of the information. For example, as Professor Scott Skinner-Thompson explained, intimate information and political information[4] "tend, by their nature, to involve higher likelihood of downstream

---

1. LYGEIA RICCIARDI, CONSUMER P'SHIP FOR EHEALTH, THE NAT'L P'SHIP FOR WOMEN & FAMILIES, PROTECTING SENSITIVE HEALTH INFORMATION IN THE CONTEXT OF HEALTH INFORMATION TECHNOLOGY 2 (June 2010), http://go.nationalpartnership.org/site/DocServer/Sensitive-Data-Final_070710_2.pdf?docID=7041 [https://perma.cc/NK5F-9BHJ].

2. Multiple studies provide empirical support for the hypothesis that consumers' assessment of the sensitivity of personal data is positively correlated with their risk perceptions of information disclosure. *See, e.g.*, Ardion Beldad et al., *I Trust Not Therefore It Must Be Risky: Determinants of the Perceived Risks of Disclosing Personal Data for E-Government Transactions*, 27 COMPUTERS HUM. BEHAV. 2233, 2237 (2011); He Li et al., *Examining Individuals' Adoption of Healthcare Wearable Devices: An Empirical Study from Privacy Calculus Perspective*, 88 INT'L J. MED. INFORMATICS 8, 8 (2016).

3. Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125, 1131 (2015); *see also* Regulation (EU) 2016/679, of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119), art. 4(1) [hereinafter *GDPR*] (defining "data subject," a common term used in the fields of data analytics, as anyone who can be identified, whether through direct or indirect means, by reference to information "such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person").

4. These refer to "sexual, medical, or mental health information" and "information arguably pertaining to countermajoritarian viewpoints," respectively. Scott Skinner-Thompson, *Outing Privacy*, 110 NW. U. L. REV. 159, 162 (2015).

consequences (such as employment discrimination resulting from the disclosed intimate information or marginalization caused by the monitoring of political thought) that they are entitled to special protection relative to other forms of information."[5] In other words, the main consideration in whether a certain type of data should receive heightened legal protection tends to be determined by the likelihood and severity of the harms that can arise from its misuse.

This association between data sensitivity and privacy risk helps to explain why both EU and U.S. laws impose heightened obligations on entities that process and control data subjects' sensitive information.[6] Several EU resolutions stipulate that special rules should govern the processing of sensitive information in view of the damage that individuals might suffer in case of misuse.[7] As early as 1995, the Data Protection Directive prohibited processing special categories of personal information; more recently, the 2018 General Data Protection Regulation (GDPR) expanded on several new conditions for this kind of sensitive data protection.[8]

Article 4 of the GDPR provides definitions for terms found throughout the regulation, which are also common in privacy and data protection discussions, such as *personal data*, *consent*, and *profiling*.[9] Notably, the GDPR makes a nuanced distinction between *controllers* and *processers*, imposing a unique set of requirements upon each.[10] A *controller* is defined as any "natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of

---

5. *Id.* ("Doctrinally, strict scrutiny is warranted because political thought and intimate information are closely related to already-recognized fundamental rights such as marital privacy, bodily integrity, and freedom of association . . . [thus, U.S. courts] appear more open to informational privacy claims when the dissemination of certain categories of information presages direct, downstream consequences, such as potential discrimination.").

6. *See* Note from the Presidency to the Council of the European Union 16525/1/12 REV 1, Data Protection Package: Report on the Progress Achieved Under the Cyprus Presidency 7 (Dec. 3, 2012), http://register.consilium.europa.eu/doc/srv?l=EN&f=ST%2016525%202012%20REV%201 [https://perma.cc/7Q2L-AWP2].

7. Éloïse Gratton, *If Personal Information Is Privacy's Gatekeeper, Then Risk of Harm the Key: A Proposed Method for Determining What Counts as Personal Information*, 24 ALB. L.J. SCI. & TECH 105, 151 (2014).

8. *See generally GDPR*, *supra* note 3, pmbl. While the 1995 Data Protection Directive needed to be transposed into Member State law, the GDPR is directly applicable to them.

9. *Id.* art. 4.

10. *See id.* art. 24–43.

personal data,"[11] while a *processor* is any such entity that processes data on behalf of a controller.[12] Under the GDPR, controllers tend to have greater responsibilities to assess and mitigate the risks of data processing,[13] defined as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means."[14] Under this expansive definition, the range of operations falling within the scope of *processing* includes collecting, recording, organizing, structuring, storing, adapting or altering, retrieving, consulting, using, disclosing by transmission, disseminating or otherwise making available, aligning or combining, restricting, erasing, and destroying data.[15]

As noted in the GDPR's recitals, the purpose of which is to express "concise reasons" for the law,[16] one of the underlying rationales for granting specific protection to sensitive information is that "[p]ersonal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms, merit specific protection as the context of their processing could create significant risks to fundamental rights and freedoms."[17]

Processing information that is sensitive in nature is also deemed riskier under U.S. law. For example, guidance on implementing the E-Government Act of 2002 indicates that the addition of health or financial information to a database "raises the risks to personal privacy" and requires the agency to conduct a Privacy Impact Assessment.[18] Other legislation and regulatory guidelines on the application of risk-based approaches to data protection further suggest accounting for the nature of personal information in assessing data processing risks.[19]

---

11. *Id.* art. 4(7).
12. *Id.* art. 4(8).
13. *Id.* art. 24.
14. *Id.* art. 4(2).
15. *Id.*
16. *See* Joint Practical Guide of the European Parliament, the Council and the Commission for Persons Involved in the Drafting of European Union Legislation 31, https://publications.europa.eu/en/publication-detail/-/publication/3879747d-7a3c-411b-a3a0-55c14e2ba732/language-en [https://perma.cc/3Y9L-8S4N].
17. *GDPR*, *supra* note 3, Recital 51.
18. *See* OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, OMB M-03-22, MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES: OMB GUIDANCE FOR IMPLEMENTING THE PRIVACY PROVISIONS OF THE E-GOVERNMENT ACT OF 2002 Attachment A, § II(B)(b)(9) (2003).
19. *See* Art. 29 Data Prot. Working Party, *Statement on the Role of a Risk-Based Approach in Data Protection Legal Frameworks* (May 30, 2014), https://www.pdp.ie/docs/10046.pdf [https://perma.cc/S4HV-X5GV]; *see also*

Finally, in both the European Union and the United States, sensitive data is thought to warrant extra protection to "avoid the risk of discrimination" based on their use by third parties.[20] Indeed, legal scholars have noted the "strong kin between data protection and discrimination issues,"[21] because data protection rights and non-discrimination rights share a common goal — guaranteeing fairness and reducing the imbalance of power between private individuals and powerful outside actors who might violate their rights.[22] In sum, the heavy obligations that come with processing sensitive information stem from the assumption that misuse of sensitive data is likely to have greater consequences for fundamental rights and freedoms than misuses of other, non-sensitive types of data.[23]

B. Legal Protections for Sensitive Data

Privacy and data protection laws and policies in both the European Union and the United States impose higher obligations upon the processing of certain types of information, referred to as *special categories of information*, *sensitive information*, or *sensitive data*. Yet, on both sides of the Atlantic, legislative bodies have taken an "ad hoc, anecdotal approach to defining sensitive information, [and]

Addendum to Note from the Presidency to the Council 10227/13 ADD 1, Subject: 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 of Such Data (General Data Protection Regulation) (May 31, 2013), https://www.huntonprivacyblog.com/wp-content/uploads/sites/18/2013/06/st10227-ad01.en13.pdf [https://perma.cc/R5S3-LNYH].

20. Yves Poullet, *About the E-Privacy Directive: Towards a Third Generation of Data Protection Legislation*, *in* DATA PROTECTION IN A PROFILED WORLD 4 (Serge Gutwirth et al. eds., 2010); *see also* Daniel L. Metayer & Julien Le Clainche, *From the Protection of Data to the Protection of Individuals: Extending the Application of Non-Discrimination Principles*, *in* EUROPEAN DATA PROTECTION: IN GOOD HEALTH? 322, 328 (Serge Gutwirth et al. eds., 2012) (arguing that data protection rights and non-discrimination rights share a common goal of guaranteeing fairness and reducing the imbalance of power between individuals and actors who may violate their rights).

21. Raphaël Gellert et al., *A Comparative Analysis of Anti-Discrimination and Data Protection Legislations*, *in* DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY: DATA MINING AND PROFILING IN LARGE DATABASES 61 (Bart Custers et al. eds., 2013).

22. *See* Metayer & Le Clainche, *supra* note 20, at 322; Poullet, *supra* note 20, at 4.

23. *See generally* PETER P. SWIRE & KENESA AHMAD, FOUNDATIONS OF INFORMATION PRIVACY AND DATA PROTECTION: A SURVEY OF GLOBAL CONCEPTS, LAWS AND PRACTICES 68 (2012); Art. 29 Data Prot. Working Party, Advice Paper on Special Categories of Data ("Sensitive Data"), Ref. Ares (2011) 444105, 4 (Apr. 8, 2011), https://www.pdpjournals.com/docs/88417.pdf [https://perma.cc/3YJ5-Y3SA] [hereinafter *Sensitive Data*].

the categories they define as sensitive change very slowly and infrequently."[24] Moreover, significant discrepancies exist between U.S. and EU law in terms of the categories of information that are considered sensitive. For example, although U.S. courts are unlikely to recognize political opinions or beliefs as a type of sensitive data due, at least in part, to First Amendment concerns, such data is explicitly protected under European law.[25] Nonetheless, as the ensuing sections of this Article seek to demonstrate, both EU and U.S. privacy and data protection laws assume that certain categories of information carry specific risks.

The following sections describe particular provisions within EU and U.S. law that protect special categories of information. In the European Union, these include Convention 108, the 1995 Data Protection Directive, and the General Data Protection Regulation (GDPR). In the United States, these include the Health Insurance Portability and Accountability Act (HIPAA), Family Education Rights and Privacy Act (FERPA), Children's Online Privacy Protection Act (COPPA), Gramm-Leach-Bliley Act (GLBA), and the Fair Credit Reporting Act (FCRA).

1. Special Categories of Information in EU Law

The European Union imposes heightened legal obligations on processing sensitive data based on what has been called a "categorical" classification scheme, meaning that certain types of information are always treated as sensitive.[26] This idea — that sensitive information is distinct *in kind* from other kinds of information, and therefore requires extra protections — was expressed in the European Union's earliest data protection laws.[27] Indeed, special legal protections for sensitive information processing have been in effect since Convention 108 first "ritualised" this by

---

24. Ohm, *supra* note 3, at 1141. Two famous examples from U.S. law include the 1998 Video Privacy Protection Act (VPPA) and the 1994 Driver's Privacy Protection Act (DPPA). It is generally accepted that the VPPA was created "almost entirely because a reporter obtained the video rental records for Judge Robert Bork during his confirmation hearings regarding his doomed nomination to the Supreme Court," while the DPPA was "inspired directly by the murder of actress Rebecca Schaeffer, killed by a deranged fan who located her using records he purchased from the California DMV." *Id.* at 1140–41.

25. *GDPR*, *supra* note 3, art. 9.

26. *See* EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, HANDBOOK ON EUROPEAN DATA PROTECTION LAW 86 (2014).

27. *Sensitive Data*, *supra* note 23, at 8.

reference to sensitive data in 1981,[28] thereby enshrining *sensitivity* as a "pivotal element of all further regulations regarding the use of personal data."[29] This Convention prohibited processing information that reveals a person's race, political and religious beliefs, health, sexual life, or criminal records in the absence of appropriate legal protection, which the Member States had to provide through law.[30]

Both the 1995 Data Protection Directive ("the Directive") and the GDPR, which went into force on May 25, 2018, strengthened protections for the discrete list of data types referred to as *special categories of information*.[31] Two specific provisions, Article 8 of the Directive and Article 9 of the GDPR, provide protections to special categories of information. By examining each in turn, it is possible to understand the boundaries of sensitive information and the rules around its processing in the EU.

The 1995 Data Protection Directive served as the fundamental framework for data protection in the European Union for over two decades, establishing the groundwork for rules controlling sensitive information processing.[32] In Article 8, the Directive allowed Member States to prohibit the processing of *special categories of personal data*, defined as information that may reveal "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and . . . data concerning health, and sex life."[33] Article 8 also contained a list of exceptions to the prohibition, such as explicit consent, necessity, protecting vital and legitimate interests of

---

28. SPIROS SIMITIS, COUNCIL OF EUROPE, REVISITING SENSITIVE DATA: REVIEW OF THE ANSWERS TO THE QUESTIONNAIRE OF THE CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING OF PERSONAL DATA 1, 1–2 (1999). In line with Convention 108, British, Dutch, and Spanish laws each grant sensitive data "special status." *Id.* Although some EU countries, including Austria and Germany, have resisted "abstract categorizations of personal data," such as *religious information* or *philosophical information*, by promoting a context-oriented appreciation for data sensitivity, they were forced to surrender this position in the face of the 1995 Data Protection Directive's explicit list of sensitive information, which was transposed into the laws of the Member States. *Id.* at 2.

29. *Id.* at 1.

30. Council of Europe, *Explanatory Report to the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data*, art. 6 (Jan. 28, 1981), https://rm.coe.int/16800ca434 [https://perma.cc/GE5F-R3UE].

31. *See* EUROPEAN UNION AGENCY FOR FUNDAMENTAL HUMAN RIGHTS, *supra* note 26, at 36.

32. Council Directive 95/46, art. 8(1), 1995 O.J. (L 281) (EC) [hereinafter *Data Protection Directive*].

33. *Id.*

individuals, and public interest.[34] As Professor Spiros Simitis argues, Member States constantly challenged the exhaustive categories of sensitive data in the Directive by "attempts to either bypass or to review the apparently definite list,"[35] resulting in a "virtually endless list of exceptions" that undermined the protections for sensitive information.[36]

The idea of granting special protection to sensitive information was strengthened when the GDPR, which repealed the Directive, came into force in mid-2018.[37] Like Convention 108 and the Directive, Article 9 of the GDPR prohibits processing special categories of information. Article 9(1) prohibits processing personal data that would reveal a person's "racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership."[38] The same provision goes on to prohibit processing "genetic data [or] biometric data for the purpose of uniquely identifying a natural person," and provides that processing of "data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited."[39]

The GDPR builds on the foundation of the laws that came before it by defining and including protections for new types of sensitive information,[40] such as data concerning health,[41] genetic data,[42] and biometric data that uniquely identifies a person,[43] and recognizes the sensitivity of information regarding sexual orientation.[44] The GDPR

---

34. *Id.* art. 8(5).
35. SIMITIS, *supra* note 28, at 3.
36. *Id.* at 3–4.
37. *Id.*; *GDPR*, *supra* note 3, art. 94(1).
38. *GDPR*, *supra* note 3, art. 9(1).
39. *Id.*
40. *Id.*
41. The GDPR defines the phrase *concerning health* as "personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status." *Id.* art. 4(15).
42. *Genetic data* is defined as "personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question." *Id.* art. 4(13). Regarding DNA information, the European Court of Human Rights has suggested that it reveals information about ethnic origin, which renders it "sensitive." S. & Marper v. United Kingdom, 2008-V Eur. Ct. H.R. 167, 196.
43. *Biometric data* is defined as "personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data." *GDPR*, *supra* note 3, art. 4(14).
44. *Id.* art. 9(1).

also introduces measures to protect children's information.[45] For example, information services geared specifically to children under the age of sixteen are not lawful without parental consent,[46] given that children may be less cognizant of the risks of data processing and their respective rights.[47] The GDPR further requires special attention from supervisory authorities and data controllers toward online services or advertisements directed towards children.[48] Data controllers are responsible for showing "reasonable efforts to verify" parental consent.[49] The need to provide transparency when processing children's information is also included in the GDPR where, as a rule of general applicability, it requires that procedures regarding the use of information should be written in "clear and plain language," a point reemphasized in the context of processing children's information.[50]

Although the GDPR is directly applicable to all EU Member States, it also provides room for the States to adopt more stringent rules on processing special categories of data.[51] Thus, Member States can introduce "further conditions, including limitations" on the processing of "genetic data, biometric data or data concerning health."[52] Member States may also create laws that lower the age requirements on processing children's data, on the condition that the age not be below thirteen years.[53]

While the EU's omnibus laws, such as the GDPR, apply to all entities that process data, whether public or private, the United States has adopted a "sectoral" approach towards data protection and privacy law — different laws regulate specific industries or uses of technology.[54] The following section examines comparable U.S. laws and how they subject certain types of information processing to heightened obligations.

---

45. *Id.* Recital 38.
46. *Id.* art. 8(1).
47. *Id.* Recital 38.
48. *Id.* art. 57(1)(b).
49. *Id.* art. 8(2).
50. *Id.* Recital 58.
51. *Id.* Recital 10.
52. *Id.* art. 9(4).
53. *Id.* art. 8(1).
54. *See* Paul M. Schwartz & Daniel J. Solove, *Reconciling Personal Information in the United States and European Union*, 102 CALIF. L. REV. 877, 879–908 (2014).

2. Information Subject to Heightened Obligations Under U.S. Law

The categories of information subjected to heightened obligations in the United States "are often [categories] which would be considered as being of 'intimate' nature," such as personal health information.[55] In contrast to EU law, nowhere in U.S. privacy law can there be found an explicit list of categories of sensitive information. Rather, "narrow sectoral laws" in the United States are targeted at specific industries.[56] Thus, determining what information qualifies as sensitive data under U.S. law can only be done in a roundabout way through an examination of the various sectoral laws. Examining various sectoral laws can help identify sensitive types of data based on the sectors that have been regulated, as U.S. privacy laws impose higher obligations on processing certain types of information in certain contexts. For instance, U.S. law subjects the processing and use of financial information, students' educational information, health information, and information about children to detailed regulations. To understand how U.S. law delineates the boundaries of sensitive information, this section briefly examines these sector-specific federal laws: HIPAA, FERPA, COPPA, GLBA, and FCRA.

Health information, which concerns the "inner workings of one's body or mind," is accorded higher protection under U.S. law because invasions of privacy in this realm may violate one's "individual sense of self."[57] Recognizing a growing need to protect private health data "in the face of digital distribution of health information,"[58] Congress enacted HIPAA.

HIPAA is a federal law that provides nationwide protection for "individually identifiable health information."[59] The Privacy Rule of HIPAA aims to ensure individuals' rights to control all forms of their health information (oral, written, and electronic) by regulating the ways other entities access and use this information.[60] It imposes

---

55. *See, e.g.*, Gratton, *supra* note 7, at 165; Skinner-Thompson, *supra* note 4, at 162.

56. Schwartz & Solove, *supra* note 54, at 881.

57. SWIRE & AHMAD, *supra* note 23, at 67.

58. Will Thomas DeVries, *Protecting Privacy in the Digital Age*, 18 BERKELEY TECH. L.J. 283, 302–03 (2003).

59. Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, § 1177, 110 Stat. 1936, 2029 (1996).

60. *See Your Rights Under HIPAA*, U.S. DEP'T OF HEALTH & HUM. SERV. (2017), https://www.hhs.gov/hipaa/for-individuals/guidance-materials-for-consumers/index.html [https://perma.cc/SYA9-NSPQ].

standardized rules on "covered entities," which include health plans, healthcare providers, and healthcare clearing houses.[61] The Privacy Rule grants individuals "the right to inspect, review, and receive a copy of their medical records and billing records that are held by health plans and health care providers" covered by the rule.[62] In addition, HIPAA's Security Rule requires covered entities to consider the probability and severity of potential risks to the confidentiality, integrity, and availability of protected electronic health information.

Students' educational records also receive special protection in the United States. The Family Educational Rights and Privacy Act (FERPA) prohibits federally-funded institutions from disclosing students' education records without parents' or eligible students' consent.[63] The law defines education records as "records, files, documents, and other materials" that involve information "directly related to a student" and "maintained by an educational agency or institution or by a person acting for such agency or institution" involved in the description of education records.[64] More specifically, education records include information about grades, attendance, disciplinary actions and course lists, as well as health and immunization records.[65] Health information about students held by a university clinic, for example, would fall under FERPA's protection of "treatment records" but excluded from coverage under the HIPAA Privacy Rule.[66]

Like the EU, the U.S. privacy framework recognizes the importance of protecting children's information. The fundamental aim of the Children's Online Privacy Protection Act of 1998

---

61. *See Summary of the HIPAA Privacy Rule*, U.S. DEP'T OF HEALTH & HUM. SERV. (2013), https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html [https://perma.cc/LY9J-SLPA].

62. *See Your Medical Records*, U.S. DEP'T OF HEALTH & HUM. SERV. (2017), https://www.hhs.gov/hipaa/for-individuals/medical-records/index.html [https://perma.cc/GWW4-A3LF].

63. *See* 20 U.S.C. § 1232(g) (2018).

64. *Id.* § 1232g(a)(4)(A).

65. *See Questions and Answers About Education Records*, U.S. DEP'T OF EDUC., https://www2.ed.gov/about/overview/focus/daca-education-records.pdf [https://perma.cc/BX2F-ERDB].

66. U.S. DEP'T OF HEALTH & HUM. SERV. & U.S. DEP'T OF EDUC., JOINT GUIDANCE ON THE APPLICATION OF THE FAMILY EDUCATIONAL RIGHTS AND PRIVACY ACT (FERPA) AND THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT OF 1996 (HIPAA) TO STUDENT HEALTH RECORDS 2 (2008), https://www2.ed.gov/policy/gen/guid/fpco/doc/ferpa-hipaa-guidance.pdf [https://perma.cc/5FC4-WA6X].

(COPPA) is to protect children from the risks of data collection through websites or mobile apps, and to engage parents in the process.[67] COPPA subjects website operators and online service providers directed at children under the age of thirteen, as well as general audience websites and online services (such as mobile apps) that knowingly collect personal information from children under the age of thirteen, to specific rules such as posting detailed privacy notices.[68] In addition, operators are required to notify parents and obtain their permission before collecting or sharing certain information regarding their children.[69]

Personal information is broadly defined under COPPA: names, addresses, online contact information, user names that include contact information (such as an e-mail address), Social Security Numbers, "persistent identifiers" that allow recognition across different services, geo-location information that allows the identification of the child's street name or city, and any "photograph, video, or audio file, where such file contains a child's image or voice."[70] It is of note that states can expand these protections. California, for example, recognizes a child's limited right to be forgotten — minors maintain the right to request that certain content be removed from websites, social networking sites, mobile apps, and other online services.[71]

Another industry subject to privacy and data protection regulations is the financial sector. The Gramm-Leach-Bliley Act (GLBA) applies to U.S. companies that are significantly engaged "in financial activities" and requires them to implement varying security programs that contain safeguards depending on the size, complexity, nature, and scope of their activities.[72] These programs should identify and assess risks to consumer information and evaluate the effectiveness of

---

67. *See Complying with COPPA: Frequently Asked Questions*, FED. TRADE COMM'N (2015), https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions#General%20Questions [https://perma.cc/G5XZ-7BDH].

68. *See* 15 U.S.C. § 6502(A)(1) (2012).

69. *Id.*

70. *See Complying with COPPA: Frequently Asked Questions*, *supra* note 67.

71. *See* CAL. BUS. & PROF. CODE §§ 22580–81 (2019).

72. 15 U.S.C. § 6801 (2012); *see also* U.S. GOV'T ACCOUNTABILITY OFF., GAO-06-674, PERSONAL INFORMATION: KEY FEDERAL PRIVACY LAWS DO NOT REQUIRE INFORMATION RESELLERS TO SAFEGUARD ALL SENSITIVE DATA (2006), https://www.govinfo.gov/content/pkg/GAOREPORTS-GAO-06-674/html/GAOREPORTS-GAO-06-674.htm [https://perma.cc/T24N-4SGQ].

current safeguards for controlling these risks.[73] The GBLA's privacy protection applies to "nonpublic personal information," which it specifically defines as "personally identifiable financial information."[74] The effectiveness of the GLBA in protecting non-public financial information, however, has been heavily criticized.[75] For example, the Act enables entities to share what might be considered sensitive information with affiliates and non-affiliates by utilizing notice and opt-out mechanisms.[76]

The last U.S. law discussed here is the FCRA, which provides protection for a very particular type of financial information: consumer reports.[77] Because a person's credit rating in the United States influences his or her loan eligibility, interest rates, or ability to rent a home,[78] consumer reports are afforded special protection under U.S. law. The FCRA is limited in scope — it applies only to consumer reporting agencies directly involved in creating consumer reports that will be used to evaluate individuals for the purposes of employment or credit.[79] Its main aim is to promote the accuracy and privacy of information regarding consumers.[80]

As the preceding sections demonstrate, laws in both the European Union and United States single out certain types of information or sectors as "sensitive," which are thereby deserving of heightened legal protections. These include health information, information about children and students, and financial information. Thus, both legal systems rely on determinations about the *nature*, *type*, or *category* of information to assess its sensitivity and riskiness. However, as Part II argues, this approach is beset by several challenges that it will be unable to address. Although prioritizing sensitive data lies at the core of laws that protect individual privacy, this approach is being rendered ineffective by innovative techniques of data collection,

---

73. *See* 15 U.S.C. § 6801.

74. 15 U.S.C. § 6802 (2012).

75. *See, e.g.*, Edward J. Janger & Paul M. Schwartz, *The Gramm-Leach-Bliley Act, Information Privacy, and the Limits of Default Rules*, 86 MINN. L. REV. 1219, 1230 (2002) ("The GLB Act has managed to disappoint both industry leaders and privacy advocates alike.").

76. *See* §§ 6802(a)–(b), 6802(b)(2).

77. *See* 15 U.S.C. § 1681 (2012).

78. *See* CHRIS JAY HOOFNAGLE, FEDERAL TRADE COMMISSION PRIVACY LAW AND POLICY 268 (2016) ("Financial privacy is an important topic because of Americans' deep dependence on the credit system.").

79. *See Credit Reporting*, FED. TRADE COMM'N, https://www.ftc.gov/tips-advice/business-center/privacy-and-security/credit-reporting [https://perma.cc/2WK7-VP7Q].

80. *Id.*

processing, and analysis. In an era characterized by new and unexpected uses of data, privacy and data protection laws are lagging years behind.

The next Part addresses three main challenges to the current approach of laws that prioritize sensitive data by category. This Article maintains that these challenges should prompt law and policymakers to reconsider ways of protecting privacy in the era of big data. Further, certain types of data no longer exist in isolation and thus ought not be guarded in isolation. As explained in the following Parts of this Article, new technologies and innovative data analytic techniques have created novel types of data, which are as worthy of protection as the traditionally-recognized categories of sensitive data, while also blurring the lines between sensitive and non-sensitive data. Moreover, lawmakers now must contend with the challenge of accounting for the ways in which the context of use for a given piece of personal information can influence its level of sensitivity. This Article concludes that these developments have rendered the current scheme of legal protections based on type, nature, or category of information as rather weak and ineffective, which leaves data subjects exposed to significant privacy risks.

II. CHALLENGES TO PROTECTING SENSITIVE DATA BY CATEGORY

As demonstrated in the previous section, the *type*, *nature*, or *category* of data — and the accompanying assumptions about the greater riskiness of sensitive kinds of data — have been used to calibrate legal obligations for data controllers and processors in both the European Union and the United States.[81] However, there are several reasons why using categorical distinctions to prioritize data determined to be "sensitive" or "high risk" may not protect individuals from the entire spectrum of privacy risks they face today.

The following sections discuss three practical challenges to laws that prioritize sensitive information. The first challenge concerns the growth in data-collecting technologies, which have led to the emergence of new types of data.[82] This new information can range

---

81. Numerous empirical studies have found that consumers' assessments of the sensitivity of personal data are positively correlated with their risk perceptions of information disclosure. *See, e.g.*, Beldad et al., *supra* note 2, at 2233.

82. *See* EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 1, 4 (2014); *see also* Liran Einav & Jonathan D. Levin, *The Data Revolution and Economic Analysis* 3 (Nat'l Bureau of Econ. Research, Working Paper No. 19035, 2013) [hereinafter BIG DATA] (explaining that "data is

from behavioral data produced by online activity and collected by social networking sites, to geolocation data produced by cell phone and other smart devices that monitor users' movements and activities.[83] These developments in data require lawmakers to constantly reevaluate and expand the list of information that can be considered "sensitive," complicating how these data can be properly protected.

A second, related challenge is that seemingly-innocuous, non-sensitive, and non-identifiable types of data can be linked to an identity, or to other sensitive data, using innovative analytic techniques.[84] Many datasets that anonymize or de-identify data — that is, tools that remove personally-identifying information such as names, addresses, or dates of birth from existing data — may be combined with other datasets to re-identify individuals.[85] The rise in the number of large, publicly-available datasets, even those that do not themselves contain any identifying or sensitive data, presents a risk to informational privacy.

A third obstacle is that the "sensitivity" of a given piece of information can change depending on the context of its use — the same piece of information might be sensitive in one context but not in another. In other words, the sensitivity of data might not be a function solely of its nature or type, but also of the way in which it is utilized or the ends to which it is put.[86] With this concern, policymakers must reconsider the effectiveness of the longstanding approach that prioritizes information according to *nature*, *type*, or *category*. Rather, policymakers must take into account *the context of data use* when crafting privacy and data protection laws.

---

now available faster, has greater coverage and scope, and includes new types of observations and measurements that previously were not available").

83. *See generally* Nicole Perlroth & Nick Bilton, *Mobile Apps Take Data Without Permission*, N.Y. TIMES (Feb. 15, 2012), https://bits.blogs.nytimes.com/2012/02/15/google-and-mobile-apps-take-data-books-without-permission/ [https://perma.cc/Y3LV-BKEM].

84. *See* BIG DATA, *supra* note 82, at 8 ("When data is initially linked to an individual or device, some privacy-protective technology seeks to remove this linkage, or 'de-identify' personally identifiable information — but equally effective techniques exist to pull the pieces back together through 're-identification.'").

85. *See id.*

86. ÉLOÏSE GRATTON, UNDERSTANDING PERSONAL INFORMATION: MANAGING PRIVACY RISKS 413 (2013).

A. Ubiquitous Collection and the Ever-Expanding Categories of Sensitive Data

The rate at which personal information is being shared, stored, and analyzed has reached unprecedented levels, ever-expanding in scope and magnitude. Today, toys "converse" with children,[87] contact lenses analyze glucose levels in tears,[88] and clothes embedded with smart devices respond to touch commands.[89] The sheer amount of data that is being generated also continues to increase in ways that defy imagination. In 2016, Dropbox users uploaded over 833,000 files, Instagram users liked nearly 2.5 million posts, Netflix subscribers streamed over 86,000 hours of video, and over 3.5 million text messages were sent in the United States alone, *for every minute of every day*.[90] Moreover, new forms of data about people, such as real-time data on their movements, preferences, and behaviors, are being collected by a growing number of interconnected devices: "The declining cost of collection, storage, and processing of data, combined with new sources of data like sensors, cameras, geospatial and other observational technologies, means that we live in a world of near-ubiquitous data collection."[91]

Individuals are also producing more and more data about themselves throughout the course of their daily activities. People share both sensitive and non-sensitive information with software and

---

87. *Cayla* is a toy that can "talk and interact, . . . play games, share photos, read stories . . . [and] can answer almost any question." *This Is Cayla*, MYFRIENDCAYLA.COM, https://www.myfriendcayla.com/meet-cayla-c8hw [https://perma.cc/AQ89-LRQ4].

88. *See* Jonah Comstock, *Novartis CEO Comments, New Patent Shed Light on Google's Contact Lens Projects*, MOBIHEALTHNEWS (Sept. 8, 2015), https://www.mobihealthnews.com/46600/novartis-ceo-comments-new-patent-shed-light-on-googles-contact-lens-projects [https://perma.cc/L3EY-J4DR]; *Digital Contact Lenses Can Transform Diabetes Care*, MED. FUTURIST, http://medicalfuturist.com/googles-amazing-digital-contact-lens-can-transform-diabetes-care/ [https://perma.cc/W9AJ-Q87U] (noting that, with the use of embedded sensors and wireless antenna communicating the information to external devices, blood glucose levels in tears will be analyzed and the data transmitted to an associated app, which will notify users to act or contact their doctors according to the results).

89. *See* David Pierce, *Google Is Hacking Our Clothes to Work Like Touchscreens*, WIRED (May 29, 2015), https://www.wired.com/2015/05/google-wants-turn-everything-wearable/ [https://perma.cc/82Z2-HLBT] ("Google is working on an ecosystem of apps and services that will let you interact with your phone and other gadgets just by grabbing, tapping, swiping, and touching your clothes.").

90. *Data Never Sleeps 4.0*, DOMO (2016), https://web-assets.domo.com/blog/wp-content/uploads/2016/06/16_domo_data-never-sleeps-4-2.png [https://perma.cc/8AJU-XPLV].

91. BIG DATA, *supra* note 82, at 4.

application providers when using their smart devices.[92] Given the real and perceived benefits accrued by using these devices and services, people voluntarily disclose sensitive information, such as data regarding their medication intake, dietary habits, or levels of fitness to app providers.[93] People who use health and fitness apps, for example, continuously provide information in real-time about their movements,[94] exercise habits and physical activity levels,[95] how much water they drink,[96] as well as how anxious or stressed[97] they feel at various times throughout the day. People who use online dating apps typically share their location, photos, and information about their hobbies and interests, with the intention of finding an agreeable partner nearby.[98] Apps used for mobile banking and other financial services collect sensitive financial information, such as Social Security Numbers, account numbers, and salary information, as well as various kinds of biometric behavioral data.[99]

It is of note that websites and app providers may collect more sensitive information than they need to merely function. For example, one of the most-downloaded gaming apps in recent years,

---

92. *See generally* Tien Wang et al., *Intention to Disclose Personal Information Via Mobile Applications: A Privacy Calculus Perspective*, 36 INT'L J. INFO. MGMT. 531 (2016).

93. One study shows that two-thirds of Americans favor using these digital tools to manage their health. *See Two-Thirds of Americans in Favor of Digital Personal Health Management*, IMAGING TECH. NEWS (Feb. 24, 2015), https://www.itnonline.com/content/two-thirds-americans-favor-digital-personal-health-management [https://perma.cc/TN53-GBDJ].

94. *See, e.g.*, *Steps — Activity Tracker*, ITUNES, https://itunes.apple.com/us/app/steps-pedometer-step-counter/id708359518?mt=8 [https://perma.cc/YW8H-DLDS].

95. *See Daily Yoga*, GOOGLEPLAY, https://play.google.com/store/apps/details?id=com.dailyyoga.inc&hl=en [https://perma.cc/JSK4-8FP7].

96. *See Water Alert*, ITUNES, https://itunes.apple.com/us/app/water-alert-drinking-water/id787142696?mt=8 [https://perma.cc/65ZL-JRMD].

97. *See Self-Help for Anxiety Management*, ITUNES, https://itunes.apple.com/us/app/self-help-for-anxiety-management/id666767947?mt=8 [https://perma.cc/KD7P-ZTYM]; *see also* Kristen Fischer, *Best Anxiety Apps of 2018*, HEALTHLINE (Apr. 30, 2018), http://www.healthline.com/health/anxiety/top-iphone-android-apps# [https://perma.cc/H7JM-ME4Z].

98. For example, one of the most popular dating apps, *Tinder*, suggests nearby "matches." *See Tinder*, GOOGLEPLAY, https://play.google.com/store/apps/details?id=com.tinder&hl=en [https://perma.cc/MMS2-SZX9].

99. *See generally* Stacy Cowley, *Banks and Retailers Are Tracking How You Type, Swipe and Tap*, N.Y. TIMES (Aug. 13, 2018), https://www.nytimes.com/2018/08/13/business/behavioral-biometrics-banks-security.html [https://perma.cc/BQ4P-KDH8].

Pokémon Go,[100] attracted criticisms for accessing users' email accounts, photos, stored data, and login information.[101] Similar outrage occurred when it was discovered that the Uber app was continuing to collect information about its users' location by running in the background even while those users were not actively using the app.[102] Social networking sites also serve as a repository for various types of sensitive data—they not only gather information that users voluntarily provide, such as their date of birth, where they live, and their employment and education history, but also generate novel and unique forms of data about users, such as information about the composition of and changes to their social networks.[103]

These technologies and platforms, which generate and collect novel forms of data, are often exempt from the heightened legal obligations placed on entities that process sensitive data.[104] Take, for example, data about social networking activities, which can be readily associated with other types of information considered to be sensitive. In one study that analyzed the "Facebook Like," which is "a mechanism used by Facebook users to express their positive association with . . . online content," researchers developed a predictive model that managed to discover sensitive information

---

100. *See* Mike Sonders, *Pokémon Go Daily Revenue: On the Decline, but There's Still Good News*, MEDIUM (Dec. 7, 2016), https://medium.com/@sm_app_intel/pok%C3%A9mon-go-daily-revenue-on-the-decline-but-theres-still-good-news-9f9b9b2b8d7 [https://perma.cc/T6E7-2P3T] ("For a couple of months after its launch, when looking at combined revenue across both iOS and Android U.S. smartphones, Pokémon GO was the top-grossing mobile game by a clear margin.").

101. *See* Laura Hudson, *How to Protect Privacy While Using Pokémon Go and Other Apps*, N.Y. TIMES (July 12, 2016), https://www.nytimes.com/2016/07/14/technology/personaltech/how-to-protect-privacy-while-using-pokemon-go-and-other-apps.html [https://perma.cc/EW98-JBSE].

102. *See* Jennifer Abel, *EPIC Fail for Uber's New Privacy Policy: FTC Asked to Block "Deceptive Data Collection*," CONSUMER AFF. (June 23, 2015), https://www.consumeraffairs.com/news/epic-fail-for-ubers-new-privacy-policy-ftc-asked-to-block-deceptive-data-collection-062315.html [https://perma.cc/JU43-37C4].

103. *See* Susan B. Barnes, *A Privacy Paradox: Social Networking in the United States*, FIRST MONDAY, Sept. 4, 2006, at 11, http://firstmonday.org/article/view/1394/1312 [https://perma.cc/2DAK-D5VH].

104. Although a variety of mobile apps and devices collect, process, and access sensitive information, they are not necessarily subject to U.S. sectoral privacy laws. For instance, apps that track people's sleep schedules, exercise habits, diets, and levels of water intake are not subject to HIPAA, which only applies to doctors, hospitals, insurers, and their business associates. *See generally* Ohm, *supra* note 3, at 1131 (arguing that HIPAA "should be expanded to include any company possessing sensitive health information," including the developers of mobile apps).

about users, including their sexual orientation, ethnicity, and religious and political beliefs.[105] To test the accuracy of the model, its predictions were compared with sensitive information that the study participants provided voluntarily.[106] In the end, the researchers were able to develop a predictive algorithm that could make "guesses" about a person's sensitive data with a high degree of accuracy, between 85% and 95%, merely from the Facebook Like data collected about individuals.[107]

The digital "identifiers" in a predictive profile, which may be linked to a particular IP address, device, or browser, are another type of data that does not fall into one of the traditionally-protected categories of sensitive data.[108] There are at least two basic types of profiles companies can create about an individual: predictive and explicit. An *explicit profile* is created when a user registers with a website and provides personally-identifying information; a *predictive profile* is created through "inference from observing and collecting user behavior over time, particularly by monitoring visited pages and ads viewed or clicked on."[109] A predictive profile can be made explicit at a later point in time — when a user (about whom a predictive profile already exists) creates an account on a website, the predictive profile can be matched with the personally-identifying information the user provides to create an explicit profile.[110] Furthermore, platforms, technology companies, and publishers have the ability to cross-match data and create individual profiles using these identifiers, and can thus follow a person's activities across smart devices.[111] While practical for users, cross-device tracking gives data

---

105. Michal Kosinki et al., *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L ACAD. SCI. 5802, 5802 (2013).

106. *Id.*

107. *Id.*

108. Poullet, *supra* note 20, at 11–13.

109. *See* GRATTON, *supra* note 86, at 412–13.

110. *See* Art. 29 Data Prot. Working Party, *Opinion 2/2010 on Online Behavioral Advertising*, 00909/10/EN, WP 171, 7 (June 22, 2010), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp171_en.pdf [https://perma.cc/8MYM-2RVY] ("The profile based on analysis of the cookies stored on the terminal equipment of the data subject can be enriched with aggregated data derived from the behavior of data subjects who exhibit similar behavioral patterns in other contexts.").

111. FED. TRADE COMM'N, CROSS-DEVICE TRACKING: AN FTC STAFF REPORT 1–2 (2017), https://www.ftc.gov/system/files/documents/reports/cross-device-tracking-federal-trade-commission-staff-report-january-2017/ftc_cross-device_tracking_report_1-23-17.pdf [https://perma.cc/8X4K-VY8S].

collectors a bigger window into users' lives.[112] A single smart device effectively stores data on one's health, finances, and location in the same place, and each smart device may be linked with any number of other smart devices.

Another example of a new form of data that can be used to identify an individual is web-browsing history, or the record of webpages a person visits over the Internet. Researchers in one study were able to identify individuals with a high degree of accuracy by combining their browsing histories with auxiliary public data.[113] As the authors of that study noted, "browsing histories contain tell-tale marks of identity."[114] Because a person is more likely to click on links shared by their specific contacts in a social network, in more than seven out of ten cases, the researchers were able to identify the "owner" of a browsing history by comparing it with the accounts that person follows on social media.[115]

The reality that web browsing history can reveal personal information about users matches anecdotes and survey evidence that suggest individuals consider their web browsing history to be a type of sensitive data.[116] A person's browsing histories can contain sensitive information about his or her health, including any stigmatizing medical conditions they might have. After being counseled by a health care professional about a sexually-transmitted disease, for example, many people seek health information related to their medical condition on the Internet.[117] Several studies have found that most Internet users disapprove of advertisers having access to records of their online behavior in order to market products and target advertisements to them.[118] This further suggests that most people

---

112. *Id.*

113. Jessica Su et al., *De-Anonymizing Web Browsing Data with Social Networks*, 26 PROC. INT'L WORLD WIDE WEB CONF. 1261, 1268 (2017).

114. *Id.*

115. *Id.*

116. *See* Ori Heffetz & Katrina Ligett, *Privacy and Data-Based Research*, 28 J. ECON. PERSP. 75 (2014); *see, e.g.*, 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 [https://perma.cc/9EQW-LKSA] (interviewing an AOL customer who was shocked to hear that AOL had saved three months' worth of her web search history).

117. *See generally* Veronique Verhoeven et al., *Everything You Always Wanted to Know About HPV (but Could Not Ask Your Doctor)*, 81 PATIENT EDUC. & COUNSELING 101 (2010).

118. *See* Chris Jay Hoofnagle et al., *Privacy and Modern Advertising: Most U.S. Internet Users Want "Do Not Track" to Stop Collection of Data About Their Online Activities*, 2012 AMSTERDAM PRIVACY CONFERENCE 1, 2 (2012); JOSEPH TUROW ET

consider their web searches to be private, sensitive information and would approve of some legal protections to that effect.

A final example of data that should be fully afforded legal protections as a category of sensitive information is data about movements and locations generated through mobile device tracking. Examining how a person's physical location changes throughout the day can reveal a detailed profile of that person's life. Several states in the United States have already decided that this sort of "tracking" warrants protection.[119] Additionally, the Supreme Court in *United States v. Jones* understood that sensitive information can be readily inferred from real-time tracking of a person's movements and location:

> Disclosed in [GPS] data . . . will be trips the indisputably private nature of which takes little imagination to conjure: trips to the psychiatrist, the plastic surgeon, the abortion clinic, the AIDS treatment center, the strip club, the criminal defense attorney, the by-the-hour motel, the union meeting, the mosque, synagogue or church, the gay bar and on and on.[120]

Because of advances in surveillance technology, this type of tracking data provides a wide window into a person's life, revealing sensitive information about him or her that simply would not have been possible to obtain, or would be prohibitively expensive to do so, just a couple of decades ago.[121] The underlying justification for giving additional protection to this type of information is that "the incremental privacy threat posed by the government's acquisition of information increases as more information is obtained . . . ."[122] As the D.C. Circuit Court explained in *United States v. Maynard*:

AL., AMERICANS REJECT TAILORED ADVERTISING AND THREE ACTIVITIES THAT ENABLE IT 3 (2009) ("Contrary to what many marketers claim, most adult Americans (66%) do not want marketers to tailor advertisements to their interests."); Lymari Morales, *U.S. Internet Users Ready to Limit Online Tracking for Ads*, GALLUP (Dec. 21, 2010), http://www.gallup.com/poll/145337/Internet-Users-Ready-Limit-Online-Tracking-Ads.aspx [https://perma.cc/MUM6-LE8G].

119. California, New Hampshire, and Maine require law enforcement officials to have a warrant for cell phone location tracking. *See* CAL. PENAL CODE § 1546 (2015); 16 ME. REV. STAT. ANN. tit. 16, § 648 (2017); N.H. REV. STAT. ANN. § 644–A (2009).

120. United States v. Jones, 565 U.S. 400, 415 (2012) (quoting People v. Weaver, 909 N.E.2d 1195, 1199 (N.Y. 2009)).

121. *Id.*

122. Matthew B. Krugler & Lior Jacob Strahilevitz, *Actual Expectations of Privacy, Fourth Amendment Doctrine, and the Mosaic Theory*, 2015 SUP. CT. REV. 205, 205 (2015); *see also* Orin S. Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311, 315 (2012).

> Repeated visits to a church, a gym, a bar, or a bookie tell a story not told by any single visit, as does one's not visiting any of these places over the course of a month. The sequence of a person's movements can reveal still more; a single trip to a gynecologist's office tells little about a woman, but that trip followed a few weeks later by a visit to a baby supply store tells a different story.[123]

Moreover, as with other types of data, location data can be combined with other datasets to identify specific individuals: "All kinds of information can be connected to a geographic location, such as financial data, health data and other consumer behavioural data."[124] Even away from personal devices, new ways of tracking people are being developed. For instance, billboards with small cameras can identify passers-by, and roadside billboards will soon use mobile location information to better target advertisements to drivers.[125]

This past section has discussed the proliferation of novel forms of data that tend to be exempt from existing laws and lists of "sensitive" information. Across the United States and the European Union, these developments will require lawmakers to constantly reevaluate and expand the list of information that is deemed worthy of heightened legal protection. This is one of several challenges to the design of privacy and data protection laws that can withstand the test of time. Turning to another key obstacle, the following section describes how the combination of "non-sensitive" pieces of information and other kinds of publicly-available data can be used to uniquely identify individuals or reveal sensitive information about them. It also shows how the proliferation of massive, open datasets, combined with advanced analytics, blurs the distinction between sensitive and non-sensitive data that has long been the heart of law and policymaking in privacy and data protection.

---

123. United States v. Maynard, 615 F.3d 544, 562 (D.C. Cir. 2010).

124. Art. 29 Data Prot. Working Party, *Opinion on Geolocation Services on Smart Mobile Devices*, 81/11/EN, WP 185, 3 (May 16, 2011), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2011/wp185_en.pdf [https://perma.cc/SJ22-LQPW].

125. *See* Sydney Ember, *See that Billboard? It May See You, Too*, N.Y. TIMES (Feb. 28, 2016), https://www.nytimes.com/2016/02/29/business/media/see-that-billboard-it-may-see-you-too.html?_r=0 [https://perma.cc/MEJ2-M7PG].

B. Uncovering Sensitive Data Through Re-Identification

Various types of data that have tended not to be considered identifiable or sensitive in isolation can, in practice, be used to uniquely identify a person, to reveal sensitive information about them, and lead to significant privacy harms. Indeed, the boundaries between different categories of information are becoming increasingly blurred today. To many, the notion that we should assign the label of "personally-identifiable information" to some data points and not to others has become "outdated."[126] Modern technologies and data analytic techniques allow sensitive information about individuals to be discerned from publicly-available or "open" datasets[127] that continuously increase in size and number.[128] Examples of these types of datasets range from user-generated reviews on websites such as Google or Yelp (which link a person to certain geographic locations, travel destinations, tastes, and financial means), to educational records from massive open online courses released by providers such as edX.[129] The expanding volume of data that exists across various channels and platforms and has become accessible to the public is thus challenging pre-determined categories of sensitive data.[130]

Researchers have shown that seemingly unrelated datasets can be aggregated to link publicly-available data about a person to that person's identity or to sensitive information about him or her. Using a method known as *data linking* or *data fusion*, which brings together data from various sources,[131] any piece of information can become identifying when combined with other bits of information.[132] As

---

126. George R. Milne et al., *Information Sensitivity Typology: Mapping the Degree and Type of Risk Consumers Perceive in Personal Data Sharing*, 51 J. CONSUMER AFF. 133, 134–36 (2016) ("[T]he PII/non-PII distinction is meaningless anyway, given the ability of current reidentification algorithms to link seemingly innocuous data into something personally identifiable.").

127. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 155 (2013).

128. *See, e.g.*, Jon P. Daries et al., *Privacy, Anonymity, and Big Data in the Social Sciences*, 57 COMM. ACM 56, 58 (2014) (pointing out that "[a]s with open source code and openly licensed content, support for open data has been steadily building" because of its "tremendous potential across the scientific disciplines to facilitate greater transparency through replication and faster innovation through novel analyses").

129. *Id.* at 59.

130. *See* Gratton, *supra* note 7, at 164.

131. *See* BIG DATA, *supra* note 82, at 4.

132. *See* Arvind Narayanan & Vitaly Shmatikov, *Privacy and Security: Myths and Fallacies of "Personally Identifiable Information"*, 53 COMM. ACM 24, 26 (2010); *see*

Professors Paul Schwartz and Daniel Solove explain, "technologists can take information that appears on its face to be non-identifiable and turn it into identifiable data."[133] In other words, the anonymity of the owners of sensitive information can be unmasked through techniques that re-identify people in so-called de-identified datasets.[134] As a 2014 White House Report on big data pointed out, "personally identifiable information can be derived or inferred from datasets that do not even include personal identifiers."[135] Anticipating these developments over twenty years ago, Joel Reidenberg and Paul Schwartz noted:

> The ability of information technology to combine and share data makes impossible any abstract noncontextual, evaluation of the impact of disclosing a given piece of personal information. The impact of bureaucratic use of personal information, whether merely personal or highly sensitive, depends on the means of processing, the kinds of databases linked together, and the ends to which information will be used.[136]

One of the most widely-cited examples of data-linking and re-identification comes from a study of the Netflix Prize. Netflix publicly released a dataset with nearly 1.5 million ratings given by 500,000 of its subscribers — this was to crowdsource improvements to the algorithm Netflix uses to predict how users rate films, and thus

---

generally Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010).

133. 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, 1816 (2011).

134. *See* BIG DATA, *supra* note 82, at 8 ("When data is initially linked to an individual or device, some privacy-protective technology seeks to remove this linkage, or 'de-identify' personally identifiable information — but equally effective techniques exist to pull the pieces back together through 're-identification.'"). For conceptual definitions of and distinctions between *anonymous data*, *explicit identifier*, and *de-identified data*, see Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* 6–7 (Carnegie Mellon Univ., Data Privacy Working Paper No. 3, 2000). According to Professor Latanya Sweeney, "[a] common incorrect belief is that removing all explicit identifiers such as name, address and phone number from the data renders the result anonymous." *Id.* For the definitional distinction between *anonymity* and *privacy*, see Jeffrey M. Skopek, *Reasonable Expectations of Anonymity*, 101 VA. L. REV. 691, 693–94 (2015) ("Although both anonymity and privacy prevent others from gaining access to a piece of personal information, they do so in opposite ways: Privacy involves hiding the *information*, whereas anonymity involves hiding what makes it *personal*.").

135. BIG DATA, *supra* note 82, at 8.

136. JOEL R. REIDENBERG & PAUL M. SCHWARTZ, DATA PROTECTION LAW AND ONLINE SERVICES: REGULATORY RESPONSES 9 (1998), http://www.paulschwartz.net/pdf/onlinesvcs_schwartz-reidenberg.pdf [https://perma.cc/PHZ4-VRS6].

better tailor film suggestions to users.[137] Although names and account names were removed, the dataset included information on all the films each user watched, the ratings they gave, and the date each rating was given.[138] Comparing the ratings in the Netflix data with ratings that were posted by known users on the Internet Movie Database (IMDb), researchers were able to specifically identify several individuals in the Netflix dataset and have access to their viewing histories.[139]

The practice of de-identifying data by removing explicit identifiers such as names, addresses, or phone numbers has been shown to be "not sufficient to render data anonymous because combinations of attributes often combine uniquely to re-identify individuals."[140] Another well-known study re-identified people in a de-identified hospital dataset that contained their diagnoses, procedures, and medications using publicly-available voter registration lists.[141] The study concluded that 87% of the U.S. population could likely be identified by their unique combination of zip code, gender, and date of birth, while more than half (53%) could be identified by their unique combination of city/town/municipality, gender, and date of birth.[142]

Data that is frequently collected by social networking sites can also be combined with ancillary data to identify or reveal sensitive information about people. In one study, researchers used facial recognition technology to link an image of a person's face to personal information about the same person that could be found online, including that person's Social Security Number.[143] Researchers were able to uncover the identity of an anonymous or unidentified person, and even retrieve sensitive information about the person in real time, simply by using social networking profiles, search queries to data aggregation websites such as Spokeo.com, statistical analysis, data mining, and facial recognition techniques.[144] It is particularly of note that the researchers were able to identify even those persons who

---

137. *See* THE NETFLIX PRIZE RULES, http://www.netflixprize.com/assets/rules.pdf [https://perma.cc/JS66-VEAA].

138. *See id.*

139. *See* Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, 2008 IEEE SYMP. ON SECURITY & PRIVACY 111, 121 (2008).

140. *See* Sweeney, *supra* note 134, at 2.

141. *See id.*

142. *See id.*

143. *See* Alessandro Acquisti et al., *Face Recognition and Privacy in the Age of Augmented Reality*, 6 J. PRIVACY & CONFIDENTIALITY 1, 10–11 (2014).

144. *See id.* at 10.

attempted to maintain anonymity online, by, for example, using a pseudonym on a dating website, as well as passers-by in the "offline" world.[145] Studies such as these demonstrate the potential of technology, much of which is readily available on the market, to pinpoint and collect sensitive information about people.

This Article has thus far examined how advancements in technology and the proliferation of big data have rendered the traditional "categorical" distinction between sensitive and non-sensitive data as outdated, and a practically ineffective tool for law and policymaking. This is because innocuous-seeming bits of information can be linked with publicly-available datasets to reveal private, personal, sensitive information. The final part of this section addresses another way in which the "categorical" approach to prioritizing sensitive data falls short of providing effective privacy protections: the fact that the sensitivity of a given piece of data is contingent upon the context in which it is used. Without acknowledging the role that context plays in data sensitivity, lawmakers in both the European Union and United States will continue to struggle to properly protect data subjects' privacy interests.

C. How the Context of Data Use Affects Sensitivity

Another practical challenge to the prioritization of sensitive data involves the relationship between the context or purpose of information use and its level of sensitivity. Relying on *type*, *nature*, or *category* of information as a guide for determining data sensitivity promotes the view that all pieces of information that fall under a given category are equally sensitive at all times — but this may not always be the case.[146] The level of risk associated with any piece of data is not entirely dependent upon its nature or type, but is usually the product of several variables: "situation-specific circumstances, the intentions of the parties involved, the kind of information being sought and the way it is processed."[147] In other words, personal

---

145. *See id.* at 7.

146. *See* Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 119, 119 (2004).

147. Gratton, *supra* note 7, at 146 n.212 ("Other relevant variables may include the historical context, the particular type of technology at stake, the political environment, the nature of the information within a given context, and the vulnerability of the individual . . . . [Also,] the long-term as well as the short-term impact on the individual affected, on what terms the information is shared, the terms of further dissemination, the purpose of disclosure, the expectations of the individual,

information may become more or less sensitive, in terms of the potential harm it can cause to individuals, depending on the context.[148] Simply prioritizing certain "categories" of information without taking into account the context of use might be too blunt an approach to protect data subjects' privacy. This approach may not only inhibit benign or beneficial uses of sensitive data, but also fail to recognize the harmful uses of data that is deemed non-sensitive in different contexts.

Indeed, numerous experimental studies have demonstrated that public perceptions about information sensitivity are shaped by the context in which information is disclosed.[149] For instance, demographic information such as name, email, mobile phone number, and mailing address are typically perceived as low in sensitivity.[150] But in the purchasing context, when demographic information is combined with financial information such as credit card and pin numbers, the same demographic information is perceived as highly sensitive.[151] Likewise, in job hunting contexts, when demographic information is combined with personal identification information such as identification photos, it is perceived to be highly sensitive.[152]

Moreover, a wide range of sensitivity levels can be found within the same "category" of data. For instance, the term *health data* may refer to a diagnosis of a common illness, such as a cold or the flu, as well as to a serious disease associated with stigmatization,[153] but to classify all information about one's health as of the same sensitivity level would be a potentially dangerous oversimplification of the data.[154]

---

the identity of the recipient, whether the recipient has an interest in knowing the information disclosed, etc.").

148. *Id.* at 164 n.318, 181 (noting that the appearance of an individual's name on a company Intranet page has fewer privacy implications than the appearance of the same name on a "blacklist" related to credit ratings. In the context of consumer privacy, the sale of one's entire consumer history (including information of an "intimate" nature) would be fundamentally more harmful than a telemarketing call based on newspaper subscription records.).

149. *See, e.g.*, David L. Mothersbaugh et al., *Disclosure Antecedents in an Online Service Context: The Role of Sensitivity of Information*, 15 J. SERV. RES. 76, 90–91 (2012).

150. *Id.* at 94.

151. *See id.*; Shu Yang & Kanliang Wang, *The Influence of Information Sensitivity Compensation on Privacy Concern and Behavioral Intention*, 40 ACM SIGMIS 38, 38 (2009).

152. Yang & Wang, *supra* note 151, at 38.

153. *Sensitive Data*, *supra* note 23, at 8.

154. *Cf. id.*

As noted earlier, a key factor in determining whether any given piece of information is sensitive is the level of risk associated with it.[155] It follows that before any piece of information can be deemed sensitive, it must be established that its exposure would be harmful to the data subject. But demonstrating the existence of such harm has been an elusive task, which has proven difficult for a variety of reasons, some judicial or constitutional and others empirical.[156]

Professor Éloïse Gratton distinguishes between three different *types of data use* by data controllers and processors: positive (uses that benefit the data subject), negative (uses that harm the data subject), and neutral (uses that neither benefit nor harm the data subject, but typically benefit the entity handling or analyzing the data).[157] Importantly, Gratton's distinctions are not between the *category*, *nature*, or *type* of information itself, but between different *uses* of information.[158] More specifically, she distinguishes between uses that create benefits, lead to harms, or do not affect the data subject at all.[159] From this standpoint, the category of the data remains constant; it is the uses or the ends towards which the data is put that alter its sensitivity level.[160] This view thus acknowledges that the same piece of information can lead to harms if used in some ways but lead to benefits if used in others, or might even remain neutral in some circumstances.

Consider, for example, different uses of data within the industry of web analytics, which seeks to understand and monetize user's online behaviors. Potential positive benefits for data subjects include more personalized services, products, and advertisements, while uses that may be neutral include those that allow an organization to develop new tools or services.[161] Yet, there are several potential harms stemming from the web data analytics:

> One could claim that providing only certain customers with specials may trigger a risk of objective harm to other customers (discrimination). Others may find that, as with targeted advertising, this may limit the various choices offered to consumers. This means

---

155. *See, e.g.*, Beldad et al., *supra* note 2 (defining sensitive health information by the amount of risk it carries in the event of disclosure).

156. *See* Skinner-Thompson, *supra* note 4, at 161 (explaining why it is so difficult to establish the existence of informational privacy harms in the courts — because privacy is conceptually tied to the complex notions of *dignity* and *autonomy*).

157. GRATTON, *supra* note 86, at 416.

158. *Id.*

159. *Id.*

160. *Id.*

161. *Id.*

> that whether information used for analytic purposes is considered personal [or sensitive] will depend on the exact use and whether this specific use may cause objective harm to an individual.[162]

If data controllers automatically or by default prioritize certain types or categories of data without considering how that data is used, they likely ignore how the benefits and risks of the same piece of data can differ from one context to another. In other words, the likelihood and severity of the harms or benefits of data processing depend upon how data is used, as well as on the context in which it is processed.[163] This feature of data processing and sensitivity-variability poses a critical challenge for lawmakers who seek to design laws that can both protect privacy and enhance beneficial data uses. Challenging as this may be, however, it is vital to contend with in the face of a world in which data is growing ever-more complex and ubiquitous.

III. RETHINKING SENSITIVE INFORMATION

This Article has thus far analyzed how policymakers in both the United States and the European Union prioritize sensitive information as a category, and discussed several practical challenges with this approach. As the border between sensitive and non-sensitive information becomes increasingly blurry due to big data, data linking, and re-identification capabilities, laws that prioritize sensitive information based on category might leave data subjects vulnerable to significant risks. Without some degree of protections for non-sensitive data categories, many privacy threats will remain — given the rapid pace of change in the data and informational landscape, risk schemes around information sensitivity need to be re-evaluated.

The practical challenges in prioritizing sensitive information are, in some ways, attributable to changes in technology. Dealing with technology that is constantly evolving has long been a key challenge for data protection laws.[164] Companies and governments retain vast amounts of information on people by connecting different data sets

---

162. *Id.* at 417.

163. *See* Nissenbaum, *supra* note 146, at 137–38 ("[W]hether a particular action is determined a violation of privacy is a function of several variables, including the nature of the situation, or context; the nature of the information in relation to that context; the roles of agents receiving information; their relationships to information subjects; on what terms the information is shared by the subject; and the terms of further dissemination.").

164. *See, e.g.*, Spiros Simitis, *Privacy—An Endless Debate?*, 98 CALIF. L. REV. 1989, 2000 (2010) (discussing Europe's attempts to pass omnibus laws as a "means to secure both a broad and reliable way to regulate the use of personal data").

and data points to create digital dossiers.[165] Today, there is no escape from leaving traces that can be recorded.[166] Moreover, because re-identification can be achieved through publicly-available tools, measures aimed at enhancing privacy, such as blurring facial images in databases or relying on individual opt-ins, may be ineffective.[167] Although laws that regulate the collection of a single type of sensitive data about users may be effective at mitigating well-known privacy risks, the collection of this data *en masse* is generating new risks.

The problems associated with prioritizing sensitive data categorically are exacerbated by the fact that people share and generate an increasing amount of data through various applications, platforms, technological accessories, and wearable devices — all of which track multiple aspects of users' daily activities.[168] Reliance on self-tracking as a means of self-improvement or self-reflection, and the concomitant desire to "quantify" oneself,[169] may continue to grow into the future. New technologies will inevitably be developed to meet this demand, producing new types and greater quantities of data, at least some of which will entail a high degree of risk and thus deserve heightened legal protection. These trends have, and certainly will continue, to fundamentally alter the boundaries of what information people consider sensitive and private.

The lists of special information in the European Union and information subject to heightened obligations under U.S. law, both discussed in this Article, are notable not only for what they include but also for what they leave out. Search queries, web browsing history, and contacts on social networking sites are examples of types of data that have tended not to be regarded as highly sensitive or risky, and have received less protection than other categories of

---

165. *See* Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1393–96, 1400–09 (2001) (highlighting the skepticism with which databases have historically been viewed, and explaining the interaction between public and private databases, the marketplace for information, and the individualized targeting that "cyberspace" affords).

166. Patrick Breyer, *Telecommunications Data Retention and Human Rights: The Compatibility of Blanket Traffic Data Retention with the ECHR*, 11 EUR. L.J. 365, 370 (2005).

167. Acquisti et al., *supra* note 143, at 14 ("Blurring of facial images in databases, k-anonymization of photos, or opt-ins, are ineffectual tools when re-identification can be achieved through already publicly available data.").

168. *See, e.g.*, DEBORAH LUPTON, THE QUANTIFIED SELF: A SOCIOLOGY OF SELF-TRACKING 2 (2016).

169. *See generally id.*

information.[170] But, as demonstrated here, this "non-sensitive data" can be used to identify or reveal sensitive information about a person, such as their health conditions, sexual preferences, political associations, and religious practices.[171]

An unfathomable number of data points can be pooled together to compile a profile for an individual within seconds. These advancements in technology have generated novel types of data, which call into question the distinction between sensitive and non-sensitive information.[172] The increasing ease with which disaggregated bits of data from around the web can be swept up and compiled into a single profile, linking presumably non-sensitive data with sensitive data, should prompt legislators to recognize these profiles as a new type of sensitive information deserving of stronger protections. Adding these novel data types to the list of special categories of information, however, would only be a first step.

As opposed to the categorical approach, the contextual approach to data considers how the context in which information is used affects its level of sensitivity.[173] This perspective shifts the focus away from the *category* of data, avoiding the question of what categories are or are not sensitive, to the *manner of data use* and its eventual consequences.[174] Many risks may not be identified if the *context* of disclosure is overlooked.[175] Uses of data that benefit the data subject, regardless of the type of data in question, should not be subject to processing restrictions, while uses of data that result in harms to the data subject often should be.[176]

Although it expanded the European Union's list of categories of sensitive information, the GDPR indicates a shift towards a

---

170. *See* Ohm, *supra* note 3, at 1142–44 ("[T]he Cable Privacy Protection Act singles out subscription information in ways that seem overprotective when compared to the fact that search queries and web history tend not to be protected.").

171. *See* Narayanan & Shmatikov, *supra* note 139, at 123.

172. *See id.*

173. Nissenbaum, *supra* note 146, at 155.

174. Gratton, *supra* note 7, at 207 (advocating that information only qualifies as personal if the way in which it is being used or collected — its context — creates a risk of harm).

175. *Id.* at 163.

176. *See generally id.* Although Gratton's framework sorts information into personal and non-personal types, rather than sensitive and non-sensitive types, the logic is essentially the same: based on the notion that the processing of some types of information (i.e., personal or sensitive) should be subject to heightened regulations, she argues that the context of data use (whether positive, negative, or neutral to the data subject) is the determining factor in whether any given information processing is deemed to involve personal information and, thus, ought to be considered risky.

contextual approach to determine information sensitivity.[177] In their advice paper, issued a year after the proposals for the GDPR were made public, the Article 29 Working Party noted that several Member States "believe that the context and/or the purpose of processing should be taken into account when assessing the issue of sensitivity."[178] The Working Party also noted that one of the shortcomings of the current categorizing approach, as embodied in the 1995 Directive, is that a "closed list [of sensitive information] is inflexible and unable to react to the context of processing as well as new forms of processing which might occur in the course of ongoing technological developments . . . ."[179]

The issue of defining and protecting sensitive information has also been subject to debate recently in the United States, especially after the Federal Communications Commission (FCC) announced a new rule that Internet Service Providers, such as Comcast and AT&T, need to require opt-in consent to collect and use *sensitive information*, including "precise geo-location, financial information, health information, children's information, social security numbers, web browsing history, app usage history, and the content of communications."[180] Broadband firms challenged the FCC's authority to impose this rule on them,[181] and President Donald Trump signed a resolution to repeal the rule.[182] Resistance by industry to regulations that would impose restrictions on the collection and use of sensitive information may be another factor that prompts lawmakers to rethink how to afford legal protections to these types of information.

---

177. *See, e.g.*, *GDPR*, *supra* note 3, Recital 76.

178. *Sensitive Data*, *supra* note 23, at 10.

179. *Id.* at 13.

180. Press Release, Fed. Commc'ns Comm'n, FCC Adopts Privacy Rules to Give Broadband Consumers Increased Choice, Transparency and Security for Their Personal Data (Oct. 27, 2016) (on file with author). Broadband customers were not given a choice to not share this information with their service providers or to use encryption. *See also* Natasha Duarte, *Frequently Asked Questions: The FCC's Broadband Privacy Rule*, CTR. DEM. & TECH. (Feb. 1, 2017), https://cdt.org/blog/frequently-asked-questions-the-fccs-broadband-privacy-rule/ [https://perma.cc/F8MC-CCU3].

181. *Digital Advertisers Battle over Online Privacy*, ECONOMIST (Nov. 5, 2016), http://www.economist.com/news/business/21709584-escalating-fight-over-users-data-and-targeted-ads-digital-advertisers-battle-over-online [https://perma.cc/F7PF-JYPQ].

182. A joint resolution providing for congressional disapproval under chapter 8 of title 5, United States Code, of the rule submitted by the Federal Communications Commission relating to "Protecting the Privacy of Customers of Broadband and Other Telecommunications Services." S.J. Res. 34, 115th Cong. (2017).

CONCLUSION

EU and U.S. information privacy laws have long relied on categorical distinctions to set priorities regarding legal protections, obligations, and restrictions on data processing activities. This approach, based on the assumption that information that is sensitive in nature tends to be associated with greater risks, remains at the core of privacy laws on both continents. The effectiveness of this approach, however, is steadily diminishing today as the long-standing distinction between sensitive and non-sensitive information, which was always on shaky ground, has all but collapsed. The exponential rise in the kinds and amount of data about us that is being collected, stored, and processed every day has brought about a reality in which a just a couple of innocuous data points are all that is needed to uniquely identify a person or to reveal sensitive information about them.

Classifying data according to its nature or type and prioritizing data considered sensitive is an essential first step in risk assessment — but this step is only the beginning. Relying entirely on the sensitivity level of a piece of data to determine the risks associated with it will fall short of adequately protecting data subjects' privacy. Data controllers and regulators must therefore consider other factors, in addition to the category of data or its sensitivity level, when determining the risks posed by data processing. There is certainly a need to interpret *sensitive information* broadly, and for laws to extend protection to newly-emerging types of data that can uniquely identify or reveal sensitive information about individuals. But it is also imperative for law and policymakers to go beyond the category-based regulation of sensitive information if they are to properly protect privacy. Moving forward, policymakers should prioritize sensitive information understood broadly, while also simultaneously protecting against the risks posed by harmful contexts of data use.

A worst-case scenario is that the legal prioritization of certain categories of information deemed special or sensitive may inadvertently lead to the neglect of risks associated with other types of data processing. It is not too much of a stretch to say that all data may be sensitive, or that there is a minimum level of sensitivity present in every piece of data. No data can ever be entirely disassociated from risk and harm. To believe otherwise would be to fail to look beyond the nature of data.