The Deidentification Dilemma: A Legislative and Contractual Proposal

Robert Gellman*

ABSTRACT

Deidentification is one method for protecting privacy while permitting other uses of personal information. However, deidentified data is often still capable of being reidentified. The main purpose of this article is to offer a legislative-based contractual solution for the sharing of deidentified personal information while providing protections for privacy. The legislative framework allows a data discloser and a data recipient to enter into a voluntary contract that defines responsibilities and offers remedies to aggrieved individuals.

INTRODUCTION

The goal of protecting the privacy interests of individuals often conflicts with ever-increasing demands for use of personal data to achieve potentially beneficial objectives in public health, law

A PDF version of this Note is available online at http://iplj.net/blog/archives/volumexxi/book1. Visit http://iplj.net/blog/archives for access to the IPLJ archive.

* Privacy and Information Policy Consultant, Washington, D.C. The author is grateful to Steven Cope, Latanya Sweeney, Peter Winn, Mark Rothstein, Susan Landau, and Joel Reidenberg for their comments and assistance with this article.

enforcement, national security, anti-terrorism, fraud prevention, and research in different fields of study.[1] Conflicts over privacy can be reduced, moderated, and balanced in various ways.

Deidentification—the removal of identifiers from personal information used or disclosed for purposes unrelated to the purpose for which the information was originally obtained[2]—is one method for protecting privacy while permitting other uses of personal information.[3] However, deidentification does not always make reidentification of individuals impossible.[4] Reidentification is the linkage of deidentified personal information with an overt identifier which belongs or is assigned to a living or dead individual.[5]

This article begins with the premise that statistical, encryption, or other mathematical approaches to deidentification aimed at protecting privacy[6] fail to provide solutions to address all data types and data sharing activities.[7] These approaches still have value because they provide some degree of privacy protection, but they seldom achieve complete deidentification of data.[8] No matter how many identifiers have been removed or encrypted and no matter how much data has been coded or masked, the remaining

---

1 Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1704 (2010) ("Data can be either useful or perfectly anonymous but never both.").

2 *See Re-identification*, ELEC. INFO. PRIVACY CTR., http://epic.org/privacy/reidentification/#intro (last visited Oct. 28, 2010).

3 *See id.*

4 *See id.*

5 *See id.*

6 Latanya Sweeney, Lab. for Computer Sci., Mass. Inst. of Tech., Roundtable Discussion: Identifiability of Data at Subcomm. on Privacy & Confidentiality, Nat'l Comm. on Vital & Health Statistics (Jan. 28, 1998), [hereinafter Sweeny, Subcomm.], *available at* http://ncvhs.hhs.gov/980128tr.htm. The National Committee on Vital and Health Statistics is an advisory committee to the U.S. Department of Health and Human Services. *Id.*

7 *See* Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of "Personally Identifiable Information,"* 53 COMMC'N ACM 24, 26 (June 2010), *available at* http://userweb.cs.utexas.edu/users/shmat/shmat_cacm10.pdf.

8 *See id.*

data may still be reidentified.[9] Further, the value of data for legitimate uses, such as research, may be significantly reduced when the data is processed without identifiers which were removed to protect privacy.[10] In the absence of a technical solution to reidentification, other approaches are needed.

The solution presented here focuses on controlling reidentification and providing accountability for those who promise not to reidentify information. This article offers a legislative-based contractual solution for the sharing of deidentified personal information while providing protections for privacy. This legislative framework allows a data discloser and a data recipient to enter into a voluntary contract that defines responsibilities and offers remedies to aggrieved individuals. Additionally, this legislative approach offers (a) common standards, (b) protections for the data subjects which are likely never to appear in private contracts, (c) a framework that can be incorporated by reference in regulations, and (d) a safe harbor provision for some activities of reidentification. The proposed contractual solution can be useful whether personal information is deidentified in support of academic research or other objectives. This proposal is not a universal guarantee of privacy, nor will it work for all data exchanges. It will, however, provide another tool to support the sharing of personal data while addressing the privacy interests of the data subjects.

In this article, deidentification means that personal information has been processed in some fashion to reduce the ability to identify the individuals to whom the data refer. It does not mean that information has been anonymized to the point where reidentification is never possible.

---

9 *See id.* ("It turns out there is a wide spectrum of human characteristics that enable re-identification: consumption preferences, commercial transactions, Web browsing, search histories, and so forth."); *see also* Ohm, *supra* note 1, at 1704–05.

10 *See* Ohm, *supra* note 1, at 1753. "[E]ven modest privacy gains require almost complete destruction of the data-mining utility." *Id.* (quoting Justin Brickell & Vitaly Shmatikov, The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing, The 14th ACM SIGKDD Int'l Conference on Knowledge Discovery & Data Mining 70, 70, 76 (August 2008)).

I. THE PROBLEM

A major challenge for deidentification is the vast amount of personal information available from public and private sources in the United States and, increasingly, elsewhere around the world.[11] The more personal data that is available, the easier it can be to link deidentified data to a particular individual.[12] The commercial collection, compilation, and exploitation of personal data in the United States are extensive.[13] Sources of personal information include public records (e.g., voter registers, occupational licenses, property ownership and tax records, court records), commercial data (e.g., transaction information), and even nonidentifiable data (e.g., census data).[14] Extensive profiles of individuals and households exist in commercial records that may include the name, address, former addresses, and telephone number of a referenced individual as well as information pertaining to her educational level, home ownership, mail order buying propensity, credit card usage, income level, marital status, age, children, and lifestyle (including personal interests in activities such as gardening or sports).[15] Private Internet companies increasingly maintain health records outside the reach of health privacy laws that only protect health records held by health care providers and insurers.[16] Internet websites, including social networking sites, are recent new facilities that provide additional sources of personal information,

---

11 *See* Narayanan & Shmatikov, *supra* note 7, at 24.

12 *See* Latanya Sweeney, *Achieving k-Anonymity Privacy Protection Using Generalization and Suppression*, 10(5) INT'L J. ON UNCERTAINTY, FUZZINESS & KNOWLEDGE-BASED SYS. 571, 588 (2002) [hereinafter Sweeney, *Achieving k-Anonymity*], *available at* http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.7384&rep=rep1&type=pdf (discussing how the increase in publically available information on the internet has created the ability to build an "electronic fingerprint").

13 *See* Latanya Sweeney, *Information Explosion*, *in* CONFIDENTIALITY, DISCLOSURE, AND DATA ACCESS: THEORY AND PRACTICAL APPLICATIONS FOR STATISTICAL AGENCIES 43, 43 (P. Doyle et al. eds., Urban Inst. 2001).

14 *See* Latanya Sweeney, Computational Disclosure Control 110 (Jan. 8, 2001) (unpublished Ph.D. dissertation, Mass. Inst. of Tech.), http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/sweeney-thesis-draft.pdf [hereinafter Sweeney, CDC].

15 *See id.*

16 *See* Robert Gellman, *Personal Health Records: Why Many PHRs Threaten Privacy*, WORLD PRIVACY FORUM (Feb. 20, 2008), http://www.worldprivacyforum.org/pdf/WPF_PHR_02_20_2008fs.pdf.

including search requests, movies watched, and other activities and interests.[17] Cellular telephones now track the location of users at all times.[18] So-called "digital signage" tracks individuals in public spaces, collecting detailed information about consumers' behavior and their characteristics, like age, gender, and ethnicity.[19]

Personal information that no longer contains overt identifiers (name, identification number, e-mail address, telephone number) can still be linked with known individuals. Identity can be ascertained from simple, basic, widely available non-unique identifiers (sometimes called *quasi-identifiers*).[20] For example, Professor Latanya Sweeney, a leading academic authority on statistics, identification, and policy, estimates that 87% of Americans can be uniquely identified from their date of birth, gender, and five-digit zip code.[21] Removing, generalizing, or coding these or other non-unique identifiers may make the task of reidentification harder, but the data may still be reidentified.[22] At the same time, deidentified data sets may be less useful for research and other uses because of the difficulty of linking data sets or because the data will no longer support complete or precise conclusions.[23]

The Health Insurance Portability and Accountability Act ("HIPAA"), a federal health privacy statute, provides an example of the difficulty of achieving—or even defining—

---

17 *See, e.g.*, J.R. Raphael, *People Search Engines: They Know Your Dark Secrets . . . And Tell Anyone*, PC WORLD (Mar. 10, 2009, 11:30 PM), http://www.pcworld.com/article/161018/people_search_engines_they_know_your_dark_secretsand_tell_anyone.html.

18 *See* April A. Otterberg, *GPS Tracking Technology*, 46 B.C. L. REV. 661, 662 (2005) (discussing the privacy implications of GPS tracking).

19 *See* Pam Dixon, *The One-Way-Mirror Society: Privacy Implications of the New Digital Signage Networks*, WORLD PRIVACY FORUM (Jan. 27, 2010), http://www.worldprivacyforum.org/pdf/onewaymirrorsocietyfs.pdf.

20 *See* R. Motwani & Y. Xu, Efficient Algorithms for Masking and Finding Quasi-Identifiers, Very Large Data Bases (VLDB) Conference, Vienna, Austria (2007).

21 Sweeney, CDC, *supra* note 14, at 20.

22 *See id.* at 203 (discussing how the Scrubs Method of de-anonymizing data cannot guarantee that data will not be reidentified).

23 *See* Sweeney, *Achieving k-Anonymity*, *supra* note 12, at 573 (explaining that information, though de-identified, must remain practically useful as research material).

deidentification.[24] HIPAA's privacy rule provides that individually identifiable health information is deidentified if seventeen specific fields of data are removed or generalized.[25] The rule assumed that deidentification by this method would provide complete anonymity to the individuals concerned. Data that is deidentified according to this standard falls outside the rule's scope, and the rule allows the data to be freely disclosed to anyone or to be published.[26]

However, notwithstanding HIPAA's determination that the resulting data is deidentified, Professor Sweeney testified that there is a 0.04% chance that data deidentified under the health Act's methodology could be reidentified when the deidentified data was compared to voter registration records for a confined population.[27] Thus, if a database deidentified under HIPAA standards had one million names, then four hundred people could likely be reidentified. If other public, commercially available, Internet-based, or private records were also to be consulted, the chances of reidentification would almost certainly increase. HIPAA's deidentification process may be the most specific and detailed regulatory approach to deidentification. Yet, even HIPAA's extensive and carefully considered efforts at deidentification do not achieve complete anonymity for all data.

Other examples in which information was reidentified after it was processed to protect privacy information can be readily found.

---

[24] *See* Health Insurance Portability and Accountability Act, 42 U.S.C. § 1320d-2 (2006). HIPAA rules cover both privacy and security. *See* HHS Security and Privacy Rule, 45 C.F.R. § 164.514(b)(2) (2010) (detailing the requirements relating to use and disclosure of health information).

[25] 45 C.F.R. § 164.514(b)(2)(i) (2010). The rule has an eighteenth, catch-all, field covering "[a]ny other unique identifying number, characteristic, or code." *Id.* § 164.514(b)(2)(i)(R). In addition to removing the specified identifiers, the entity making the disclosure cannot have actual knowledge that the information "could be used alone or in combination with other information to identify an individual who is a subject of the information." *Id.* § 164.514(b)(2)(ii) (2010).

[26] *See* NAT'L COMM. ON VITAL & HEALTH STATISTICS, U.S. DEP'T OF HEALTH AND HUMAN SERVS., ENHANCED PROTECTIONS FOR USES OF HEALTH DATA: A STEWARDSHIP FRAMEWORK FOR "SECONDARY USES" OF ELECTRONICALLY COLLECTED AND TRANSMITTED HEALTH DATA 36 (2007), *available at* http://www.ncvhs.hhs.gov/071221lt.pdf.

[27] *Id.* at 36 n.16.

Professors Krish Muralidhar and Rathindra Sarathy provide a case study using educational performance data publicly released by states under federal rules.[28] The disclosures are supposed to comply with a requirement in the Family Educational Rights and Privacy Act of 1974 ("FERPA") that the disclosure of identifiable student data requires written parental consent.[29] Muralidhar and Sarathy show that even when the release of aggregate data satisfies standards for minimum cell sizes, the remaining data can still allow for the computation of personally identifiable information about particular subgroups and about individuals.[30]

A June 2010 article by Arvind Narayanan and Vitaly Shmatikov noted the shortcomings of deidentification and offered a broad and general conclusion: "[t]he emergence of powerful re-identification algorithms demonstrates not just a flaw in a specific anonymization technique(s), but the fundamental inadequacy of the entire privacy protection paradigm based on 'de-identifying' the data."[31]

Professor Paul Ohm suggests that "[u]ntil a decade ago, the robust anonymization assumption worked well for everybody involved."[32] He provides additional examples of several well-publicized releases of supposedly deidentified data that were ultimately found to be identifiable, including the America Online release of research data search queries and the Netflix release of a prize data study that contained 100 million movie ratings by Netflix customers.[33]

Indeed, there may not be a realistic and practical standard for absolute deidentification in today's data rich world. Professor Sweeney put it this way: "I can never guarantee that any release of [deidentified] data is anonymous, even though for a particular user

---

[28] Krish Muralidhar & Rathindra Sarathy, PRIVACY VIOLATIONS IN ACCOUNTABILITY DATA RELEASED TO THE PUBLIC BY STATE EDUCATIONAL AGENCIES 1 (2009), *available at* http://gatton.uky.edu/faculty/muralidhar/EdPrivacyViolation.pdf.

[29] 20 U.S.C. § 1232g(b) (2006).

[30] *See* Muralidhar & Sarathy, *supra* note 28, at 7–20.

[31] Narayanan & Shmatikov, *supra* note 7, at 26.

[32] Ohm, *supra* note 1, at 1716.

[33] *Id.* at 1720–24.

it may very well be anonymous."[34] As a general proposition, for most personal data, deidentification may be like absolute zero for temperature: a state that can be approached but never achieved. Even if data could be fully deidentified, the prize may not be worth the effort in many cases. The data may no longer have significant value for researchers and other users.[35]

From a policy perspective, identifiability of personal information is best viewed as a continuum. At one end of the continuum, information is fully identifiable due to the presence of names, identification numbers, and the like. Shedding overt identifiers moves data down the continuum where it becomes harder to link the data with individuals, but data may still be identifiable even with all overt identifiers removed. While it may be possible at times to achieve provably absolute deidentification using encryption, coding, hashing, and other techniques,[36] it seems highly unlikely that there is a general solution that will work for all types of data, all types of users, and all types of activities. Thus, we continue to face the possibility that deidentified personal data shared for research and other purposes may be subject to reidentification.

II. EXISTING LEGAL APPROACHES

Statisticians have long been aware of deidentification issues and have developed many techniques to address the possibility of reidentification.[37] However, existing laws do little to untangle the deidentification dilemma. Indeed, they tend to make it worse. Existing laws often reflect an assumption that identifiability is a binary state; personal data is either identifiable or it is not. These

---

34 *See* Sweeney, Subcomm., *supra* note 6.

35 *See* Ohm, *supra* note 1, at 1704.

36 *See* Sweeney, Subcomm., *supra* note 6.

37 *See, e.g.*, Fed. Comm. on Statistical Methodology, *Report on Statistical Disclosure Limitation Methodology* 12–28 (Statistical Policy Working Paper No. 22, 2005), *available at* http://www.fcsm.gov/working-papers/spwp22.html (discussing sampling, suppression random rounding, and swapping as examples of methods for deidentification); Khaled El Emam, *Heuristics for De-identifying Health Data*, 6 IEEE SEC. & PRIVACY 58–61 (2008) (discussing heuristics that are used and how they can be applied).

laws tend to ignore the reidentification issue altogether, establishing vague or inconsistent standards for identifiability. Some examples:

- The Privacy Act of 1974,[38] a U.S. law that applies mostly to federal agencies, defines "record" as a grouping of information about an individual that contains "his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph."[39] An identifier is an essential part of a record.[40] The ability to infer identity or to reidentify a record is not sufficient or relevant, no matter how easy it may be to accomplish the reidentification. Further, the law treats a fingerprint as an identifier,[41] when few people without access to a law enforcement fingerprint database could identify an individual from a fingerprint. The Act's concept of identifiability is muddled, at best.

- The Cable Communications Policy Act does not define "personally identifiable information," but it excludes from the term "any record of aggregate data which does not identify particular persons."[42] However, even aggregate data can be used to reidentify individuals in some circumstances.[43] The statute does not address that possibility.

---

38 5 U.S.C. § 552a (2006).

39 *Id.*

40 *See* 5 U.S.C. § 552a(a)(4) (referring to examples of identifiers such as name, identifying number, symbol, etc.).

41 *See id.*

42 47 U.S.C. § 551(a)(2)(A).

43 *See* Muralidhar & Sarathy, *supra* note 28, at 7–20 (discussing how the disclosure of personally identifiable information results from the release of aggregate data regarding individual education performance).

- The Confidential Information Protection and Statistical Efficiency Act of 2002 ("CIPSEA") defines "identifiable form" as "any representation of information that permits the identity of the respondent to whom the information applies to be reasonably inferred by either direct or indirect means."[44] CIPSEA's definition is one of the few that explicitly addresses the use of indirect inferences to permit identification,[45] but it does not indicate the scope of effort that is necessary to render deidentified data identifiable. Further explication would presumably require parsing the meaning of "reasonably."

- Canada's Personal Information Protection and Electronic Documents Act ("PIPEDA") defines "personal information" as "information about an identifiable individual."[46] Thus, PIPEDA offers no standard for determining identifiability or anonymity, nor does it address the issue of reidentification. A treatise on PIPEDA suggests that truly anonymous information does not qualify for protection under the statute.[47] It also suggests that "caution should be exercised in determining what is truly 'anonymous' information since the availability of external

---

[44] Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002, 44 U.S.C. § 3501.

[45] *Compare* The Privacy Act of 1974, 5 U.S.C. § 552a, *and* Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, § 2(1) (Can.), *available at* http://www.canlii.org/en/ca/laws/stat/sc-2000-c-5/latest/sc-2000-c-5.html (defining "personal information" without explicitly addressing information that indirectly permits personal identification), *with* Confidential Information Protection and Statistical Efficiency Act of 2002, 44 U.S.C. § 3501.

[46] Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, § 2(1) (Can.), *available at* http://www.canlii.org/en/ca/laws/stat/sc-2000-c-5/latest/sc-2000-c-5.html.

[47] STEPHANIE PERRIN, ET AL., THE PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMENTS ACT: AN ANNOTATED GUIDE (2001).

information in automated format may facilitate the reidentification of information that has been made anonymous."[48] That advice may be helpful, but the statute itself is silent.

- The European Union ("E.U.") Data Protection Directive defines "personal data" as "any information relating to an identified or identifiable natural person," and it defines an identifiable person as "an individual person . . . who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."[49] The task of parsing these words for a clear standard is helped somewhat by the Directive's Introductory Recital 26, which states that privacy rules will not apply to "data rendered anonymous in such a way that the data subject is no longer identifiable."[50] It also provides that "to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person."[51] Based on the Recital, it seems apparent that the Directive uses a reasonableness standard to determine whether information is sufficiently deidentified to fall outside the Directive's ambit.

---

48 *Id*. at 54.

49 Council Directive 95/46, art. 2, Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995 O.J. (L 281) 31 (EC), *available at* http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri= CELEX:31995L0046:en:HTML.

50 *Id*. at Recital 26.

51 *Id*.

A further gloss on the Directive's meaning of "personal data" can be found in an opinion of the Article 29 Working Party, an organization established under the Directive.[52] The opinion offers twenty-six pages of detailed and interesting explanation of how to establish what is and is not personal data, and what is and is not identifiable.[53] The length of the analysis is evidence of the complexity of identifiability under current conditions. The Working Party's conclusion that a determination of anonymity depends on the circumstances and calls for a case-by-case analysis is further evidence of the essential murkiness of the identifiability concept.[54]

- The Alberta Health Information Act defines "individually identifying" to mean when a data subject "can be readily ascertained from the information,"[55] and it defines "non-identifying" to mean that the identity of the data subject "cannot be readily ascertained from the information."[56] This appears to limit the identifiability inquiry to the information itself. Alberta's data matching law[57] regulates the creation of individually identifying health information by combining individually identifying or non-identifying health information or other information from two or

---

52 *See id.* at Art. 29 (establishing the Working Party as an independent, advisory committee with representatives from each state and the European Union).

53 Article 29 Data Protection Working Party, *Opinion 4/2007 on the Concept of Personal Data*, 01248/07/EN WP 136 (June 20, 2007), *available at* http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf.

54 *See id.* at 21.

55 Health Information Act, R.S.A. 2000, c. H-5 § 1(p) (Can.), *available at* http://www.qp.alberta.ca/574.cfm?page=H05.cfm&leg_type=Acts&isbncln=9780779739493.

56 *Id*. § 1(r).

57 *Id*. § 68.

> more electronic databases without the consent of the data subjects.[58] The data matching requirements include submission of a privacy impact assessment to the commissioner for review and comment.[59] The Alberta law expressly addresses reidentification activities by anyone (at least, anyone using an electronic database).[60] The Act establishes an administrative process rather than a statutory standard for determining whether identifiable information is at stake.

In general, statutes and rules that address identifiability and deidentification can be grouped roughly into three categories.[61] The first category includes standards that seek to determine whether data is sufficiently or potentially identifiable to warrant regulation. The standards can (a) be inward-looking (considering only the data);[62] (b) be outward-looking (considering other data actually or potentially available elsewhere as well as the capabilities for reidentification generally available to individuals or experts);[63] (c) require professional statistical judgment;[64] or (d) consider the time, effort, or cost required for reidentification.[65] More than one of these standards can apply at the same time. A standard can also reference a reasonableness test, either directly or indirectly.[66] As is apparent, a multiplicity of standards are available in this category.[67]

---

58 *Id*. § 1(g).

59 *Id*. §§ 70(3), 71(3).

60 *See id.* at § 1(g) (limiting scope of data matching provisions to combinations of electronic databases).

61 *See* Robert Gellman, *Privacy for Research Data*, PUTTING PEOPLE ON THE MAP: PROTECTING CONFIDENTIALITY WITH LINKED SOCIAL-SPATIAL DATA 81, 92 (Nat'l Research Council ed., 2007), *available at* http://books.nap.edu/catalog.php?record_id=11865 (reviewing privacy laws in Australia, Britain, Canada, Europe, and the U.S.).

62 *Id.*

63 *Id.*

64 *Id*.

65 *Id*.

66 *Id.* at 88–89. The European Union Data Protection Directive is an example of a statute that employs an explicit reasonableness test. *See supra* notes 49–52 and accompanying text. The British Data Protection Act incorporates a reasonableness test

The second category uses an administrative process. The Alberta law calls for administrative privacy review in advance of some reidentification activities.[68] This type of review may be possible in a small jurisdiction, but it would be impractical in a larger one.

The third category is a rule that requires the removal of specified data elements. HIPAA's health privacy rule is a leading example.[69] Some of its shortcomings have already been discussed.[70]

This limited review of statutes and rules suggests the wide variance in identifiability standards that can be found. In most cases, the statutes offer alternate word formulas[71] that are probably casually drafted rather than a carefully considered approach based on detailed study or analysis. Recognition by legislators and policy makers of the complexities presented by deidentification of personal data has been slow to develop. Laws badly trail the capabilities of modern computers and experts to use the vast pools of personal data available today.[72] Current technology allows for the reidentification of data that most casual observers would have thought was adequately deidentified.[73]

All existing regulatory approaches suffer from shortcomings.[74] HIPAA's rule provides greater certainty, but that certainty is somewhat misplaced.[75] CIPSEA expressly recognizes the possibility of reidentification, but it offers little practical guidance.[76] It might well take years of litigation before any useful test emerges, and the result of litigation may not provide enough

implicitly. *See* UK Data Protection Act, 1998 c. 29, § 1(1), *available at* http://www.legislation.gov.uk/ukpga/1998/29/contents.

67 *See* Gellman, *supra* note 61, at 92.

68 *See id*.; *see also* Health Information Act, R.S.A. 2000, ch. H-5 §§ 68–72 (Can.).

69 *See* 45 C.F.R. § 164.514(b)(2)(i) (2010).

70 *See supra* pp. 37–38.

71 *See* Gellman, *supra* note 61, at 92.

72 Ohm, *supra* note 1, at 1706 ("[P]owerful advances in reidentification thwart the aims of nearly every privacy law and regulation.").

73 *See supra* notes 9, 22 and accompanying text.

74 Ohm, *supra* note 1, at 1706.

75 *See supra* note 27 and accompanying text.

76 *See supra* notes 44–45 and accompanying text.

clarity. The Alberta administrative process likely will not scale to larger jurisdictions or will require a cumbersome bureaucracy.[77]

III. CONTRACTUAL SOLUTION

No legislation can establish meaningful standards for the creation of deidentified data that has full value for legitimate secondary users. That objective cannot be reached now and may be impossible to achieve generally. There will always be a tradeoff of some sort, involving the degree of identifiability of the data, the usability of the data, the privacy of the data subjects, and the cost of the deidentification process.[78] Technology can sometimes lessen these tradeoffs, but it cannot eliminate them all the time.[79]

What legislation can do, however, is establish a statutory framework that will allow the data disclosers and the data recipients to agree voluntarily on externally enforceable terms that provide privacy protections for the data subjects. An ordinary contract or exchange of data is not likely to consider or give effect to the rights of the data subjects, who are rarely if ever parties to the transaction.[80] The proposed statute defines the terms of data disclosure and rights for the data subjects. The effect is to strike a balance between the interests of all parties: the data disclosers, the data users, and the data subjects. That is the main purpose of the

---

[77] *See supra* note 57–60 and accompanying text. *But see infra* Appendix, Personal Data Deidentification Act ("PDDA"). The Alberta law requires the submission of a privacy impact assessment to the commissioner for review and comment, which is probably only feasible in a smaller jurisdiction because of the level of review required. *See* Health Information Act, R.S.A. 2000, c. H-5 § 70(2). The PDDA is probably better suited to a larger jurisdiction than the Alberta law because it does not require administrative review prior to reidentification. *See generally infra* Appendix, Personal Data Deidentification Act.

[78] *See* Ohm, *supra* note 1, at 1751–52 (noting that as data utility increases, privacy protection decreases, and technology has yet to eliminate this tradeoff).

[79] *See id.*

[80] *See* Gellman, *supra* note 61, at 110. Since the parties to a data agreement are the data discloser and the data recipient, a data subject has difficulty suing for breach of contract due to a lack of privity. However, a data subject may escape the problem of lack of privity in some jurisdictions by suing as a third party beneficiary. *See id.*; *see also infra* notes 112–14 and accompanying text.

law proposed here, the Personal Data Deidentification Act ("PDDA").[81]

The key definition in the PDDA is *potentially identifiable personal information* ("PI^2"). The definition of PI^2 builds on a definition of *personal information*, which is any information about an individual, whether it contains a personal identifier or not.[82] *Potentially identifiable personal information* is any personal information without overt identifiers.[83] PI^2 is a new concept in the PDDA, included to cover the wide range of personal information without overt identifiers that is likely to be reidentifiable. Since it cannot be known at any time whether information is reidentifiable,[84] virtually all personal information that is not overtly identifiable is PI^2. Aggregate data (as opposed to microdata, which is data about an individual)[85] is not expressly addressed in the proposal, but the proposed contractual solution could work just as well for aggregate data that includes the possibility of reidentification.

The core proposal in the legislation is a voluntary *data agreement*, which is a contract between a data discloser and a data recipient.[86] The PDDA will only apply to those who choose to accept its terms and penalties through a data agreement.[87] The PDDA establishes standards for behavior and proposes civil and criminal penalties for violations.[88] The data recipients would be prohibited from reidentifying or attempting to reidentify any potentially identifiable personal information under the threat of

---

81 *See infra* Appendix, Personal Data Deidentification Act, Preamble.

82 *See infra* Appendix, Personal Data Deidentification Act §§ 2(6)–(7).

83 *See infra* Appendix, Personal Data Deidentification Act § 2(7).

84 *See* Gellman, *supra* note 61, at 85 ("It may not always be easy to predict in advance when deidentified data can be linked. . . . Whether a set of data is identifiable can depend on the characteristics of the set itself, on factors wholly external to the set, or on the identity of the observer." (citation omitted)).

85 *See Linked Social-Spatial Data: Promises and Challenges*, *in* PUTTING PEOPLE ON THE MAP: PROTECTING CONFIDENTIALITY WITH LINKED SOCIAL-SPATIAL DATA 7, 11 (Myron P. Guttman & Paul C. Stern, Nat'l Research Council eds., 2007).

86 *See infra* Appendix, Personal Data Deidentification Act § 3.

87 *See infra* Appendix, Personal Data Deidentification Act § 3(a).

88 *See infra* Appendix, Personal Data Deidentification Act §§ 4–5, 7(a)–(b) (specifying civil and criminal penalties).

civil and criminal penalties.[89] The data recipients would also be required to maintain technical, administrative, and other safeguards against reidentification.[90] In exchange, there are benefits to the discloser and recipient. Some disclosers would benefit from the proposed safe harbor provision that offers liability protections to those who apply the statute for research activities.[91] Recipients would benefit by being able to offer potential disclosers more assurance that a data transfer will not create liabilities.[92] The data subjects would benefit from uniform rules, new protections, and enforcement methods that are difficult or impossible to find today.[93]

The proposal would not require all data disclosers and all data users to comply with its requirements. Only those who voluntarily choose to reference the PDDA in a contract or equivalent document would be subject to its requirements.[94] A mandatory solution may not be practical. There appears to be no way to write a definition that would encompass all data transfers, and there are too many data transfers to expect that one size will fit all.[95]

A voluntary approach allows those who want the benefits to accept the obligations. A model for this approach to legislation is arbitration. Laws define, support, and provide for the enforcement of arbitration agreements,[96] but it is typical for the parties of a contract to decide whether they want to make use of arbitration at all.[97] If they do not, then arbitration laws do not affect their activities.[98]

---

89 *See infra* Appendix, Personal Data Deidentification Act §§ 4(1), 7.

90 *See infra* Appendix, Personal Data Deidentification Act § 4.

91 *See infra* Appendix, Personal Data Deidentification Act § 6.

92 *See infra* Appendix, Personal Data Deidentification Act § 6.

93 *See* Gellman, *supra* note 61, at 110; *see also supra* Part II.

94 *See infra* Appendix, Personal Data Deidentification Act § 3(a).

95 *See supra* notes 38–60 and accompanying text.

96 *See* Federal Arbitration Act, 9 U.S.C. § 2 (2006).

97 *See, e.g.*, Ass'n of Am. R.Rs. v. Surface Transp. Bd., 162 F.3d 101, 111 (D.C. Cir. 1998) ("It cannot be gainsaid that the submission of a dispute to arbitration is normally a voluntary act, either at the time of the dispute or at an earlier time in a contract providing for such arbitration.").

98 *See id.*

The discloser who shares data under a data agreement proceeds with the knowledge that the recipient has accepted strict limits on the data's use and disclosure, and that these restrictions are enforceable by the state and by the data subjects.[99] The discloser does not have to accept any obligation to police the agreement or to act on behalf of the data subjects, other than to report breaches of the agreement to a government agency.[100] The main benefit to the discloser is that criminal liability is capped or eliminated by the law.[101] Further, the proposal includes a formal safe harbor provision, which exempts a discloser from liability for disclosure under a data agreement if (1) the recipient is a government agency, non-profit organization, or research organization that has not reported a breach of a data agreement in the five years prior to date of the agreement,[102] and (2) the disclosure is for use in research ("systematic investigation designed to develop or contribute to generalizable knowledge, but does not include marketing research")[103] or in a public health activity.[104] The purpose of the safe harbor provision is to provide encouragement for data sharing for beneficial purposes.

The recipient who seeks data under a data agreement is in a better position to ask for data from a discloser because the data agreement and the law impose on the recipient defined and enforceable limits that protect the privacy of the data subjects.[105] A reluctant source might be encouraged by a would-be recipient to share data because of the existence of formal standards and limited liability.[106] The recipient accepts the limits and the liability as a condition of receiving the data.[107]

The data subject benefits from disclosure under a data agreement because of strong rules prohibiting conduct that could

---

99 *See infra* Appendix, Personal Data Deidentification Act § 4.
100 *See infra* Appendix, Personal Data Deidentification Act § 5.
101 *See infra* Appendix, Personal Data Deidentification Act § 7(b).
102 *See infra* Appendix, Personal Data Deidentification Act § 6(a)(1).
103 *See infra* Appendix, Personal Data Deidentification Act § 2(10).
104 *See infra* Appendix, Personal Data Deidentification Act § 6(a)(2).
105 *See infra* Appendix, Personal Data Deidentification Act § 4.
106 *See infra* Appendix, Personal Data Deidentification Act § 4.
107 *See infra* Appendix, Personal Data Deidentification Act § 3(a).

reidentify data.[108] The data subject also benefits because the law clarifies that the data subject is a third party beneficiary of the data agreement.[109] This enables an aggrieved data subject to seek damages from a negligent party to the data agreement.[110] Under current law, a data subject may be unable to sue relying upon an ordinary contract between a data discloser and a data recipient because the data subject is not a party to the contract.[111] "The data subject lacks privity–an adequate legal relationship–to the contract and cannot use the contract to enforce his or her interest."[112] Today, only in some jurisdictions will the data subject be recognized as a third party beneficiary of a data use agreement and be able to seek damages.[113] In general, however, the requirement for privity can be a major obstacle to enforcement of privacy rights by the data subjects.[114] The proposed law would clarify this issue in favor of the data subjects.

Most of the obligations fall on the data recipient, which is appropriate because the recipient obtains new data vulnerable to reidentification. The recipient agrees to:

> 1) not reidentify or attempt to reidentify any potentially identifiable personal information received under [a] data agreement;
> 2) take reasonable steps . . . to prevent any . . . related party from reidentifying or making an attempt to reidentify any potentially identifiable personal information . . . received under that data agreement;

---

108 *See infra* Appendix, Personal Data Deidentification Act § 4.

109 *See infra* Appendix, Personal Data Deidentification Act § 8(a).

110 *See infra* Appendix, Personal Data Deidentification Act § 8(b).

111 *See* Joel Reidenberg, *The Privacy Obstacle Course: Hurdling Barriers to Transnational Financial Services*, 60 FORDHAM L. REV. S137, S175 (1992) (citing Brian Napier, Contractual Solutions to the Problem of Equivalent Data Protection in Transborder Data Flows, Presentation at Conference on Legal Challenges and Opportunities Created by the Prolific Growth of Electronic Information Services (Mar. 27–28, 1990) (paper on file with the Fordham Law Review)).

112 Gellman, *supra* note 61, at 110; *see also* Reidenberg, *supra* note 111, at S175.

113 *See* Gellman, *supra* note 61, at 110.

114 *See id*.

> 3) not further use or disclose any potentially identifiable personal information received under that data agreement except in accordance with that data agreement;
> 4) only disclose potentially identifiable personal information received under that data agreement to another person if the disclosure is allowed by that data agreement and if the disclosure is made pursuant to that data agreement or another data agreement subject to [the] Act;
> 5) maintain reasonable physical, administrative, and technical safeguards to protect against reidentification of potentially identifiable personal information received under that data agreement;
> 6) inform a potential discloser in writing before entering into any data agreement . . . of any actual or reasonably likely breaches of other data agreements . . . that the recipient entered into during the past 10 years.[115]

The last requirement provides a self-policing mechanism that obliges bad actors to tell others before they can obtain new data.[116]

The fourth requirement—that data received under a data use agreement only be redisclosed under the original data agreement or under a new data agreement subject to the Act—means that there must be a chain of trust if data is further disclosed.[117] The data involved in a data use agreement will always be subject to the mandated protections.[118] If allowed by the original data use agreement, the data recipient can become a data discloser with respect to the next recipient, and the protections continue in force because a new data use agreement is required.[119]

Both the recipient and the discloser must: (1) report any breach of a data agreement that the recipient entered into to a national

---

115 *See infra* Appendix, Personal Data Deidentification Act §§ 4(1)–(6).
116 *See infra* Appendix, Personal Data Deidentification Act § 4(7).
117 *See infra* Appendix, Personal Data Deidentification Act § 4(4).
118 *See infra* Appendix, Personal Data Deidentification Act § 3(a)(1).
119 *See infra* Appendix, Personal Data Deidentification Act § 4(4).

consumer protection/privacy agency and to each other;[120] (2) publish a notice of the breach prominently on their respective public websites;[121] and (3) maintain the website notice for two years.[122] In addition, in the event that either party learns that potentially identifiable personal information under their data agreement has been reidentified, that party must comply with applicable security breach notification laws.[123] The proposed law does not impose a security breach notification obligation of its own. It references existing obligations.[124] It can be anticipated that parties to a data agreement will allocate responsibility for compliance with breach notification laws among themselves in some suitable manner.

The PDDA includes several carefully tiered criminal penalties for violations. The penalties range from civil penalties for failure to report or failure to post,[125] to felonies for knowing and willful reidentification or attempted reidentification,[126] to major felonies with the possibility of imprisonment for knowing and willful reidentification or attempted reidentification with the intent to sell, transfer, or use personal information for commercial advantage, personal gain, or malicious harm.[127] There is also a felony for disclosing PI^2 obtained under a data agreement subject to the Act in violation of the terms of the agreement.[128]

Civil remedies are available to an individual whose PI^2 has been reidentified against a discloser or recipient who is negligently responsible for the reidentification.[129] The PDDA specifically provides that a data subject is a third party beneficiary of a data agreement so that there will be no issue about a lack of standing to sue over the contract.[130]

120 *See infra* Appendix, Personal Data Deidentification Act §§ 4(7)(A), 5(1)(A).
121 *See infra* Appendix, Personal Data Deidentification Act §§ 4(7)(B), 5(1)(B).
122 *See infra* Appendix, Personal Data Deidentification Act §§ 4(7)(C), 5(1)(C).
123 *See infra* Appendix, Personal Data Deidentification Act §§ 4(8), 5(2).
124 *See infra* Appendix, Personal Data Deidentification Act §§ 4(8), 5(2).
125 *See infra* Appendix, Personal Data Deidentification Act § 7(a).
126 *See infra* Appendix, Personal Data Deidentification Act § 7(b)(2).
127 *See infra* Appendix, Personal Data Deidentification Act § 7(b)(3).
128 *See infra* Appendix, Personal Data Deidentification Act § 7(b)(1).
129 *See infra* Appendix, Personal Data Deidentification Act § 8.
130 *See infra* Appendix, Personal Data Deidentification Act § 8(a).

Other provisions require an appropriate government agency with oversight responsibilities for the Act to file a biennial report, review the law in five years, and prepare model data agreements.[131] Another provision makes it clear that the PDDA does not change, override, or preempt any requirement or obligation established by other laws, and does not exempt anyone from complying with obligations under any law or rule for the protection of human research subjects.[132] Finally, the proposed legislation has been drafted in a manner that is not directly tied to U.S. law.[133] The same approach might have value in other jurisdictions.

CONCLUSION

The proposed PDDA seeks to strike a balance between the need to share deidentified personal information for research and other purposes and the inability to guarantee that the information is wholly deidentified. The solution is to allow the data disclosers and the data recipients to enter into a voluntary data agreement that defines the obligations of the parties, provides greater certainty about the potential liabilities, and allows individual data subjects to enforce their privacy interests when data has been reidentified.[134] In order to support appropriate sharing, the legislation includes a safe harbor provision for a data discloser who shares data for a beneficial purpose.[135]

Today's lack of clear definitions, deidentification procedures, and legal certainty can impede some useful data sharing. It can also affect the privacy of users when the lack of clarity about deidentification results in sharing of identifiable data that could have been avoided.

The proposed approach to the deidentification dilemma faced by data processors and policy makers will not solve every problem associated with personal data transfers and uses, but it will make available a new tool that fairly balances the needs and interests of

131 *See infra* Appendix, Personal Data Deidentification Act § 9.
132 *See infra* Appendix, Personal Data Deidentification Act § 10.
133 *See generally infra* Appendix, Personal Data Deidentification Act.
134 *See supra* notes 86, 88–89 and accompanying text.
135 *See supra* notes 91, 102–04 and accompanying text.

the data disclosers, the data users, and the data subjects. The solution could be invoked voluntarily by the data disclosers and the data recipients. Its use could also be mandated by regulation or legislation seeking to allow broader use of personal data for beneficial purposes.

APPENDIX

A BILL

To protect the privacy of potentially identifiable personal information by establishing accountability for the use and transfer of potentially identifiable personal information. [Version 4.4]

SECTION ("SEC.") 1. SHORT TITLE.

This Act may be cited as the "Personal Data Deidentification Act."

SEC. 2. DEFINITIONS.

As used in this Act:

(1) DATA AGREEMENT.—The term "data agreement" means a contract, memorandum of understanding, data use agreement, or similar agreement between a discloser and a recipient relating to the use of personal information.

(2) DATA AGREEMENT SUBJECT TO THIS ACT.—The term "data agreement subject to this Act'' means a data agreement between a discloser and a recipient who have entered into an agreement described in section 3(a).

(3) DISCLOSER.—The term "discloser" means a person who discloses potentially identifiable personal information to another person pursuant to a data agreement subject to this Act.

(4) OVERT IDENTIFIER.—The term "overt identifier" means any personal information that identifies or can readily be used to identify a particular individual, and includes a name, address, Social Security number, account number, license number, serial number, telephone number, electronic mail address, Internet protocol address, webpage address, or biometric, that alone or in combination with other information identifies or can readily be used to identify a particular individual.

(5) PERSON.—The term "person" means an individual, corporation, company, foundation, association, society, partnership, firm, non-profit organization, school, college, or university, or a department, agency, or other instrumentality of [Federal, State, or local] government.

(6) PERSONAL INFORMATION.—The term "personal information" means information about an individual that may or may not include an overt identifier.

(7) POTENTIALLY IDENTIFIABLE PERSONAL INFORMATION.—The term "potentially identifiable personal information" means any personal information without any overt identifiers.

(8) PUBLIC WEBSITE.—The term "public website" means a facility by which a person displays information to the general public on the Internet or any comparable successor technology.

(9) RECIPIENT.—The term "recipient" means a person who receives potentially identifiable personal information from another person pursuant to a data agreement subject to this Act.

(10) RESEARCH.—The term "research" means a systematic investigation designed to develop or contribute to generalizable knowledge, but does not include marketing research.

(11) REIDENTIFICATION.—The term "reidentification" means linking potentially identifiable personal information to an overt identifier belonging or assigned to any living or dead individual.

SEC. 3. DATA AGREEMENTS.

(a) AGREEMENTS SUBJECT TO ACT.—A person who enters into a data agreement that expressly references this Act by including the words "This data agreement is subject to the Personal Information Deidentification Procedures Act" or equivalent words, or who is required to be subject to this Act by statute or regulation for any disclosure or receipt of potentially identifiable personal information

(1) shall be bound by, and subject to, all of the terms of this Act with respect to potentially identifiable personal information disclosed or received under that data agreement, statute, or regulation; and

(2) may not terminate, revoke, suspend, or otherwise limit or restrict the application of this Act to potentially identifiable personal information disclosed or received under that data agreement, statute, or regulation.

(b) ADDITIONAL TERMS PERMITTED.—The parties to a data agreement subject to this Act may include additional terms to that data agreement that do not limit or undermine the terms required by this Act.

SEC. 4. DUTIES OF RECIPIENT.

A recipient under a data agreement subject to this Act shall—

(1) not reidentify or attempt to reidentify any potentially identifiable personal information received under that data agreement;

(2) take reasonable steps, including contracts, technical measures, or workplace rules, to prevent any employee, agent, consultant, contractor, affiliate, subcontractor, or other related party from reidentifying or making an attempt to reidentify any potentially identifiable personal information that the recipient received under that data agreement;

(3) not further use or disclose any potentially identifiable personal information received under the data agreement except in accordance with that data agreement;

(4) only disclose potentially identifiable personal information received under that data agreement to another person if the disclosure is allowed by that data agreement and if the disclosure is made pursuant to that data agreement or another data agreement subject to this Act;

(5) maintain reasonable physical, administrative, and technical safeguards to protect against reidentification of potentially identifiable personal information received under that data agreement;

(6) inform a potential discloser in writing before entering into any data agreement that will be a data agreement subject to this Act with the potential discloser of any actual or reasonably likely breaches of other data agreements subject to this Act that the recipient entered into during the past 10 years;

(7) (A) promptly report any breach of a data agreement subject to this Act that the recipient entered into to—

(i) the [National Consumer Protection/Privacy Agency]; and

(ii) the discloser;

(B) promptly publish a notice of the breach prominently on the recipient's public website; and

(C) maintain the notice for two years; and

(8) in the event that the recipient learns that potentially identifiable personal information that the recipient obtained under that data agreement has been reidentified, comply with applicable [Federal or State] security breach notification laws.

SEC. 5. DUTIES OF DISCLOSER.

A discloser under a data agreement subject to this Act shall—

(1) (A) promptly report any breach of that data agreement to the [National Consumer Protection/Privacy Agency];

(B) promptly publish a notice of the breach prominently on the discloser's public website; and

(C) maintain the notice for two years;

(2) in the event that the discloser learns that any potentially identifiable personal information that the discloser disclosed under that data agreement has been reidentified, comply with applicable [Federal or State] security breach notification laws; and

(3) in the event that the discloser learns that any potentially identifiable personal information disclosed under that data agreement has been reidentified or may have been reidentified, immediately suspend further disclosures of potentially identifiable personal information to the recipient under the data agreement.

SEC. 6. SAFE HARBOR.

A discloser who lawfully discloses potentially identifiable personal information under a data agreement subject to this Act—

(1) to a recipient who is a government agency, non-profit organization, or research organization that has not reported a breach of a data agreement in the five years prior to the date of the agreement,

(2) for use in research or in a public health activity,

shall not be liable under this Act or any other law to an individual who is the subject of potentially identifiable personal information disclosed pursuant to that data agreement for any damage resulting from that disclosure, except in the case of gross negligence on the part of the discloser.

SEC. 7. PENALTIES.

(a) CIVIL PENALTIES.—A person who fails to report a breach of a data agreement subject to this Act, or to post a notice in accordance with this Act, shall be subject to a civil penalty of not more than [$2,000] in an action brought by the [National Consumer Protection/Privacy Agency][Attorney General].

(b) FELONY OFFENSES.—

(1) A recipient, or any employee, agent, consultant, contractor, affiliate, subcontractor, or other related party of a recipient, who willfully discloses potentially identifiable personal information received under a data agreement subject to this Act in violation of this Act is guilty of a felony and shall be fined not more than [ ] or imprisoned not more than [ ] years, or both.

(2) A recipient, or any employee, agent, consultant, contractor, affiliate, subcontractor, or other related party of a recipient, who willfully reidentifies or attempts to reidentify potentially identifiable personal information received under a data agreement subject to this Act in violation of this Act is guilty of a felony and shall be fined not more than [ ] or imprisoned not more than [ ] years, or both.

(3) A recipient, or any employee, agent, consultant, contractor, affiliate, subcontractor, or other related party of a recipient, who willfully reidentifies or attempts to reidentify potentially identifiable personal information received under a data agreement subject to this Act in violation of this Act with the intent to sell, transfer, or use personal information obtained under a data agreement subject to this Act for commercial advantage, personal gain, or malicious harm is guilty of a felony and shall be fined not more than [ ] or imprisoned not more than [ ] years, or both.

(c) FAILURE TO INFORM OFFENSES.—A person who fails to inform a potential discloser in writing before entering into any data agreement subject to this Act as required by section 4(8) is guilty of a misdemeanor and shall be fined not more than [ ].

SEC. 8. CIVIL REMEDIES.

(a) BENEFICIARIES OF THE AGREEMENT.—An individual who is the subject of potentially identifiable personal information that is disclosed pursuant to a data agreement subject to this Act shall be a third party beneficiary of that data agreement.

(b) LIABILITY FOR NEGLIGENCE OF DISCLOSERS AND RECIPIENTS.—If a discloser or recipient of potentially identifiable personal information pursuant to a data agreement subject to this Act fails to exercise reasonable care to prevent the reidentification of an individual who is the subject of the information, that individual may bring a civil action against the discloser or recipient if the individual suffers monetary harm, emotional harm, reputational harm, or public embarrassment as a result of such reidentification. Any individual entitled to damages under this subsection shall recover not less than $1000, and the court may award reasonable attorney fees and other reasonable litigation costs to an individual who substantially prevails.

SEC. 9. DUTIES OF [NATIONAL CONSUMER PROTECTION/PRIVACY AGENCY].

The [National Consumer Protection/Privacy Agency] shall—

(1) make a biennial report summarizing any activities under this Act [to the national legislature] and post the report on its public website;

(2) evaluate the operations of the Act and report to [the national legislature] within five years after the date of enactment of this Act; and

(3) within six months of the date of enactment, publish one or more model data agreements.

SEC. 10. OTHER LAWS.

Nothing in this Act changes, overrides, or preempts any requirement or obligation established by any other law. Nothing in this Act exempts any person from complying with obligations

under any applicable law or rule for the protection of human research subjects.