2009

Privacy by Deletion: The Need for a Global Data Deletion Principle

Benjamin J. Keele
bkeele@iu.edu

Privacy by Deletion: The Need for a Global Data Deletion Principle

BENJAMIN J. KEELE*

ABSTRACT

With global personal information flows increasing, efforts have been made to develop principles to standardize data protection regulations. However, no set of principles has yet achieved universal adoption. This note proposes a principle mandating that personal data be securely destroyed when it is no longer necessary for the purpose for which it was collected. Including a data deletion principle in future data protection standards will increase respect for individual autonomy and decrease the risk of abuse of personal data. Though data deletion is already practiced by many data controllers, including it in legal data protection mandates will further the goal of establishing an effective global data protection regime.

INTRODUCTION

With the rise of digital storage and information searching technologies, people's lives have become increasingly documented.[1] Financial transactions, visits to the doctor, job applications, and even web searches produce information that can be used to identify a particular individual and serve as a record of his or her private activities. If one could keep track of and control all of one's data, perhaps this

* Editor-in-Chief, *Indiana Journal of Global Legal Studies*. J.D. candidate, 2009, Indiana University Maurer School of Law — Bloomington, B.A. *with highest distinction*, 2006, University of Nebraska—Lincoln. Email: benjamin.j.keele@gmail.com. Many thanks to Professor Fred H. Cate and the staff of the *Indiana Journal of Global Legal Studies*. All errors are mine alone. This note is dedicated to my wife, Christy.

1. DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 1–3 (2004).

Indiana Journal of Global Legal Studies Vol. 16 #1 (Winter 2009)

ubiquitous record-keeping would cause little concern. But data does not stay in one place, or even one country. Instead, personal information is frequently transferred across national borders.[2] Multinational corporations transfer data between offices in different nations; organizations outsource data processing to entities in other countries; and citizens of different nations complete transactions with each other, transferring personal data in the process.

This increased data flow creates uncertainty about which data protection laws apply to personal information. Nations protect their citizens' information privacy in different ways, and entities that collect, process, or store personal information (data controllers) have the difficult task of determining how to comply with applicable laws. If the data controllers have trouble knowing what they may do with a given set of data in a given jurisdiction, how can the individuals whose information is in the data set—the data subjects—know what can be done with their personal information?

There are many benefits to global transfers of personal data,[3] but confusion and disagreement over treatment of personal information in different jurisdictions can prevent, and have prevented, these benefits from being fully realized. For example, European law has been an obstacle to some transfers of data, such as airline passenger data, to the United States.[4] The United States government was interested in transfers of airline passenger data for national security reasons. While an interim agreement was eventually reached that enabled the United States to receive the information and a final accord is near,[5] for a time, air travel between the United States and Europe was threatened by disagreement over how European travelers' Passenger Name Records (PNR) would be transferred, processed, and stored. Also, the European Union has taken steps against data controllers who retain data for too long; the situation of America-based search engine Google illustrates the difficulties divergent data protection laws can cause.[6] In sum, inconsistent data protection laws have impeded beneficial global transfers of

2. Miriam Wugmeister et al., *Global Solution for Cross-Border Data Transfers: Making the Case for Corporate Privacy Rules*, 38 GEO. J. INT'L L. 449, 449 (2007).

3. *Id.*

4. Ellen Nakashima, *U.S., E.U. Miss Deadline on Data-Sharing Agreement*, WASH. POST, Oct. 1, 2006, at A14. For background on the PNR disagreement, see generally Irfan Tukdi, Comment, *Transatlantic Turbulence: The Passenger Name Record Conflict*, 45 HOUS. L. REV. 587 (2008).

5. Charlie Savage, *U.S. and Europe Near Agreement on Private Data*, N.Y. TIMES, June 28, 2008, at A1.

6. Google was warned for keeping search logs for two years. *See* Kevin J. O'Brien & Thomas Crampton, *European Union Warns Google on Possible Violations of Privacy Law*, N.Y. TIMES, May 26, 2007, at C3.

personal information and have slowed important interactions between governments, companies, and individuals on a global scale.[7]

A promising solution to this quandary is global data protection principles—standards developed by international bodies that serve as models for drafting and harmonizing nations' data protection laws. If a set of data protection principles achieved widespread acceptance, then governments, multinational corporations, and other global actors would have consistent data protection standards that would greatly ameliorate the difficulties caused by the current patchwork of data protection regulations.

Unfortunately, no set of principles has yet reached this tipping point to become a universally adopted, global standard. The Organization of Economic Cooperation and Development's Guidelines on the Protection of Privacy and Transborder Flows of Personal Data[8] have influenced many later data protection laws.[9] The European Union's Council Directive on the Protection of Individuals with Regard to the Processing of Personal Data and the Free Movement of Such Data[10] (hereinafter EU Directive) has been the foundation of data protection law within the European Union. Despite the importance of these standards, however, no global consensus on data protection has been reached. Furthermore, the existing standards are vague and thus unlikely to produce uniform results even if widely adopted. Even the most specific set of principles, the EU Directive, permits wide variation in national data protection laws.[11] Most other candidates for global data protection standards, like the Asian-Pacific Economic Cooperation Framework, are voluntary and anticipate divergent implementation by national governments.[12]

This note offers a proposal for a data deletion principle to be included in future sets of privacy standards. The data deletion principle will require data con-

7. Wugmeister, *supra* note 2, at 449–50. For other examples, see FRED H. CATE, PRIVACY IN THE INFORMATION AGE 126–27 (1997).

8. Organization for Economic Cooperation and Development, *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, OEC Doc. 93200201HE5 (Oct. 1, 1980), *available at* http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html [hereinafter OECD].

9. Fred H. Cate, *The Failure of Fair Information Practice Principles*, *in* CONSUMER PROTECTION IN THE AGE OF THE 'INFORMATION ECONOMY' 348 (Janet K. Winn ed., 2006).

10. Council Directive 95/46, 1995 O.J. (L 281) 31–50 (EU), *available at* http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML [hereinafter EU Directive].

11. Andrew Charlesworth, *Information Privacy Law in the European Union: E Pluribus Unum or Ex Uno Plures?*, 54 HASTINGS L.J. 931, 939 (2003).

12. Carla Bulord, Note, *Between East and West: The APEC Privacy Framework and the Balance of International Data Flows*, 3 I/S: J. L. & POL'Y FOR INFO. SOC'Y 705, 711 (2008).

trollers, overseen by legally authorized public or private sector data protection agencies, to establish record destruction schedules that will apply consistently across national boundaries and industries. Failure to comply with one's own data destruction schedule could result in fines or liability to any individual harmed by data that were not properly destroyed.

While the idea of a data deletion principle is not entirely novel, the international consensus that has coalesced around other data protection principles has not yet encompassed a deletion principle. This note argues for a strong but flexible data deletion principle, and fleshes out the various policy considerations affected by a legal mandate to destroy personal information. Global data protection standards should cover the entire life cycle of information, up to and including the necessary destruction of personal information. Until a data deletion principle is adopted as an integral part of a data protection regime that protects privacy while permitting global data transfers, no data protection scheme will be complete.

Part I will review past sets of data protection principles and demonstrate that data deletion has not been given sufficient attention and treatment. Part II will argue that the omission of a data deletion principle is a fundamental weakness in any global data protection regime. Part III will set out a proposed data deletion principle and examine the advantages of and obstacles to its adoption.

I. THE MISSING DATA DELETION PRINCIPLE

With the advent of digital information storage and processing, personal information can be more easily collected, stored, transferred, and processed.[13] This raises important privacy concerns. Data protection principles have been developed at the national and regional levels to allow for the use of personal data in beneficial ways while reducing the risk of abuse. A data deletion principle would help achieve these goals, but unlike many data protection principles, data deletion has not yet been generally accepted.

A. Data Protection: Helping to Preserve Individual Autonomy in a Bureaucratic World and Prevent Harmful Uses of Personal Information

Privacy and data protection implicate numerous social values and civil rights, including autonomy over one's body, freedom from unjustified searches, control

13. SOLOVE, *supra* note 1, at 3–4.

over one's public persona, and privacy in one's personal thoughts.[14] Data protection principles seek to address one particular aspect of privacy: maintaining appropriate control over non-public data that relate to and identify a particular individual. Examples of such personal information include records of financial transactions, medical conditions, private online activities, and academic and employment histories. An appropriate balance must be struck between permitting economically beneficial and socially essential uses of personal data while preventing harm to data subjects and giving data subjects sufficient control to maintain a minimum degree of autonomy.[15]

Data protection principles generally regulate the collection, storage, transfer, and processing of personal data. These are important functions because as individuals' lives become more susceptible to recording and analysis, entities and institutions are more likely to use personal data to make important decisions. Personal data can also be used by entities, institutions, and third parties capable of directly or indirectly accessing the information in ways that may harm individuals identified by the data.

Many important decisions about individuals are based upon their personal information. Most applications for credit cards, car loans, or mortgages are screened on the basis of personal information contained in credit bureau reports. Many employers run criminal background checks on prospective employees. Insurers check personal data before issuing insurance policies. It is clear that many essential tasks, ranging from obtaining a checking account to purchasing medicine containing pseudoephedrine, require the collection, storage, and analysis of individuals' personal information.

Much of this personal information is processed in bureaucratic organizations, entities that are characterized by a "hierarchical chain-of-command, specialized offices to carry out particular functions, and a system of general rules to manage the organization."[16] Professor Solove notes that bureaucracies, while generally efficient and impartial, ignore the unusual needs of particular individuals; avoid accountability by obscuring decision-making processes; and fail to control abuses of functionaries' discretion.[17]

14. *See* FRED H. CATE, PRIVACY IN PERSPECTIVE 3–4 (2001).
15. *See* CATE, *supra* note 7, at 31.
16. SOLOVE, *supra* note 1, at 38.
17. *Id.* at 39.

> Bureaucratic decision-making processes are being exercised ever more frequently over a greater sphere of our lives, and we have little power or say within such a system, which tends to structure our participation along standardized ways that fail to enable us to achieve our goals, wants, and needs.[18]

Thus, while bureaucratic processing of personal data is necessary and even beneficial, it must be regulated to give individuals some measure of control over their personal information, thereby forcing data controllers to respect the individuals' autonomy and allowing the individual to participate in these important decision-making processes. Data protection principles (and, through modeling and harmonization, national data protection laws) can regulate data controllers to protect individual dignity and autonomy in the use of personal information.

Aside from allowing sufficient individual control in personal data processing, data protection principles also help reduce the risk of abuses of personal data. There are myriad ways in which personal information may be used to the severe detriment of data subjects, ranging from the financially crippling (fraud, identity theft, or insurance discrimination) to the relatively innocuous but irritating (unwanted telephone or mail solicitations).[19] In addition to causing the loss of individuals' money, time, patience, or reputation, abuse of personal information also reduces individuals' confidence in the systems that utilize personal data, thus limiting the potential social benefits of such systems. For example, widespread adoption of electronic health records, an innovation that could reduce health care costs and facilitate valuable medical research, has been slowed, in part, by disagreements over rules balancing patient privacy against beneficial uses of data.[20] Data protection principles that help reduce the risk of these harms will protect data subjects' interests and increase confidence in legitimate and useful personal information systems. These perceived benefits have prompted several attempts to develop global data protection standards.

---

18. *Id.*

19. CATE, *supra* note 14, at 6–7.

20. Milt Freudenheim & Robert Pear, *Health Hazard: Computers Spilling Your History*, N.Y. TIMES, Dec. 3, 2006, at 31.

B. Current Global Data Protection Standards: Failing to Consistently Regulate Data Deletion

Beginning in the 1970s with the United States' Fair Information Practice Principles,[21] governments have attempted to develop standards capable of serving as conceptual foundations for data protection policies. While some sets of international privacy principles have mentioned destruction of data as part of particular principles, data deletion has never been deemed to merit its own complete data protection principle. The following review of established data protection standards will illustrate accepted data protection norms and the absence of data deletion in those standards.

1. OECD Guidelines

The first global privacy standards were proposed by the Organization for Economic Cooperation and Development (OECD) in 1980. The OECD Guidelines were heavily influenced by earlier United States government work on fair information practices.[22] While not universal, the OECD Guidelines are the most widely accepted data protection standards. Thus, the OECD Guidelines represent baseline data protection standards and have served as the foundation for later principles.[23]

The OECD Guidelines contain eight principles, most of which have appeared in later sets of standards with minor variations and under slightly different names.[24] The Collection Limitation Principle[25] limits how personal data are collected. The data must be collected by lawful and fair means, and generally, data subjects should know about the data collection and provide consent. The Data Quality Principle[26] deals with how the data are maintained. Data should be kept accurate and current to the extent the purpose of the data set requires. The Purpose Specification Principle[27] mandates that the purpose for the data collection be stated no later than at the time of collection. The data cannot later be used for a purpose "incompatible" with the original collection purpose. What is "incompatible" is not entirely clear, but the

---

21. *See* SEC'Y ADVISORY COMM. ON AUTOMATED PERS. DATA SYS., U.S. DEP'T OF HEALTH, EDUC. & WELFARE, RECORDS, COMPUTERS AND THE RIGHTS OF CITIZENS (1973).

22. Cate, *supra* note 9, at 348.

23. Professor Cate notes that "most of the dozens of national and regional privacy regimes adopted after 1980 claim to reflect the OECD Guidelines." *Id.*

24. While the precise details of all data protection principles are not relevant to this Note, see generally *id.*, for a concise but detailed history of data protection principles.

25. OECD, *supra* note 8, ¶ 7.

26. *Id.* ¶ 8.

27. *Id.* ¶ 9.

Purpose Specification Principle does reduce to some extent "function creep" (data being collected for one purpose and then used for a different purpose later). For example, under this principle, information collected for assessing taxes could not later be used for issuing drivers' licenses. The Use Limitation Principle[28] supports the Purpose Specification Principle by demanding that data not be disclosed or used for purposes other than the ones specified. Exceptions are made with the data subject's consent or when authorized by law.

The Security Safeguards Principle[29] requires reasonable precautions against loss, unauthorized access, modification, disclosure, or destruction of data. The Openness Principle[30] states that individuals should be able to easily determine if a data controller has data about them and how that data may be accessed. Flowing from the Openness Principle is the Individual Participation Principle,[31] which provides that procedures should exist by which data subjects may access their data. If the data are inaccurate, the data subject should be able to challenge the data and have them corrected or deleted. This is the only mention of data deletion in the OECD Guidelines. Finally, the Accountability Principle[32] simply states that data controllers should be held accountable for complying with the other principles.

While collection, use, disclosure, and storage of personal data are included, the Guidelines lack a principle regarding the destruction of data, the last stage of the data life cycle. Only when data are incorrect does a data subject's privacy interest require deletion (if modification of such data does not correct the inaccuracy). Perhaps because the OECD Guidelines, the first and most widely accepted set of data protection principles, did not substantively deal with the issue of data deletion, later actors did not feel compelled to regulate the last stage of the data life cycle.

2. Council of Europe Convention

A year after the OECD Guidelines were adopted, the Council of Europe promulgated the Convention for the Protection of Individuals with Regard to the Collection and Processing of Personal Data on Information Highways (COE Convention).[33] The COE Convention contains principles similar to the OECD

28. *Id.* ¶ 10.
29. *Id.* ¶ 11.
30. OECD, *supra* note 8, ¶ 12.
31. *Id.* ¶ 13.
32. *Id.* ¶ 14.
33. Council of Europe, Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Jan. 28, 1981, 20 I.L.M. 317, *available at* http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm [hereinafter COE Convention].

Guidelines' Collection Limitation, Use Limitation, Purpose Specification, Data Quality, Security Safeguards, and Individual Participation Principles.[34] An interesting difference in the COE Convention's Data Quality Principle is a requirement that once the need for identifying specific individuals has passed, the data should be anonymized to avoid identifying specific individuals.[35] Thus, while the COE Convention does not require complete destruction of any data, it does require deletion of data that can identify a specific person when the specified purpose of the data set no longer requires identification of individuals. The requirement that data be deleted if inaccurate and challenged by the data subject also appears in the COE Convention.[36] Although the COE Convention furthered the notion of data deletion in its data protection standards, it ultimately fell short of creating a data deletion standard.

3. EU Directive

In 1995, the European Union (EU) adopted the Directive on the Protection of Individuals with Regard to the Processing of Personal Data and the Free Movement of Such Data (EU Directive).[37] The EU Directive, similar to its data protection predecessors, contains principles on collection limitation, purpose specification, data quality, security, openness, and individual access. Like the COE Convention, the Data Quality Principle of the EU Directive requires anonymization of data when the data's purpose has been served.[38] Additional innovations include a requirement that the amount of data collected not be excessive relative to the collection's specified purpose.[39] New principles were also enacted to restrict transfers of data only to recipients who provide an "adequate level of protection" for the data,[40] to specially protect "sensitive" data (data on racial or ethnic origin, political opinions, religious beliefs, philosophical or ethical persuasion, and health and sexual life),[41] and to give individuals the right to know the logic of any processes that use their data to make automated decisions.[42] Enforcement principles were also added: one for independent oversight of

34. *Id.* arts. 5, 7, 8.
35. *Id.* art. 5(e).
36. *Id.* art. 8(c).
37. EU Directive, *supra* note 10.
38. *Id.* art. 6(e).
39. *Id.* art. 6(c).
40. *Id.* art. 25.
41. *Id.* art. 8.
42. *Id.* art. 12. An example of an automated decision is an automated process that analyzes certain financial facts about an individual and thereby determines whether to grant the individual a loan.

data controllers by government agencies, and one giving individuals enforceable rights against data controllers who violate national data protection laws.[43]

The EU Directive has been called "the high-water mark of substantive legal protection for information privacy"[44] and, due to its transborder transfer restrictions, "the closest approximation to a strong global data protection standard in operation."[45] However, the EU Directive suffers from a major omission: a principle of data deletion. National data protection laws could regulate data deletion, but to comply with the EU Directive, national laws need only require anonymization when the purpose for collecting the personal information no longer requires identification and deletion when data are challenged and found to be incorrect.

A notable feature of the COE Convention and EU Directive is that they impose legal duties on the signatory national governments to harmonize their data protection laws with the principles.[46] Supranational authorities can, to some extent, enforce the harmonization of national laws, facilitating data transfers between signatory countries. The flip side of this coin is that non-signatory nations are barred from receiving data transfers unless they are deemed to have sufficient data protection safeguards.[47] This has necessitated the Safe Harbor Agreement between the EU and the U.S. Department of Commerce to enable transatlantic data transfers.[48] Any future global data protection standard will have to either supplant the European standards or meet their minimum requirements.

4. APEC Privacy Framework

In 2004, the Asian-Pacific Economic Cooperation (APEC) sought to modernize the OECD Guidelines. The final result was the adoption of the APEC Privacy Framework (Framework).[49] The Purpose Specification Principle disappeared in this iteration, and new principles were added. The Framework now

43. *Id.* arts. 22–24.

44. Cate, *supra* note 9, at 351.

45. Sunni Yuen, *Exporting Data with Trust: Audited Self-Regulation as a Solution to Cross-Border Data Transfer Protection Concerns in the Offshore Outsourcing Industry*, 9 COLUM. SCI. & TECH. L. REV. 41, 69 (2008), *available at* http://www.stlr.org/cite.cgi?volume=9&article=2.

46. COE Convention, *supra* note 33, art. 4; EU Directive, *supra* note 10, art. 1.

47. CATE, *supra* note 7, at 126.

48. For a summary of the Safe Harbor Agreement, see A.B.A. SECTION OF SCIENCE & TECHNOLOGY LAW, INTERNATIONAL GUIDE TO PRIVACY 94–97 (Jody R. Westby ed., 2004).

49. Asian-Pacific Economic Cooperation, APEC Privacy Framework, Nov. 20, 2004, *available at* http://www.ag.gov.au/www/agd/rwpattach.nsf/VAP/(03995EABC73F94816C2AF4AA2645824B)~APEC+Privacy+Framework.pdf/$file/APEC+Privacy+Framework.pdf [hereinafter APEC Framework].

contains the Preventing Harm Principle,[50] which focuses on reducing the risk of harmful misuses of personal data. Data controllers should have specific obligations to avoid harm and safeguards should be proportional to the risk of harm. The Notice Principle[51] is similar to the Purpose Specification Principle in that data subjects should be notified of the general privacy policies relating to the data collection. The Collection Limitation Principle[52] and Uses of Personal Information Principle[53] require that data should be collected and used only in accordance with the purposes of the data collection. However, rather than using the "incompatible" criterion found in the OECD Guidelines, there are three exceptions for using data for other purposes: when consent of the data subject is obtained; to provide a service requested by the data subject; and by authority of law.[54]

The Framework further emphasizes procedural rights. The Choice Principle[55] directs data controllers to give individuals more choices regarding the collection, use, and disclosure of data relating to the individuals. The Integrity of Personal Information Principle,[56] Security Safeguards Principle,[57] Access and Correction Principle,[58] and Accountability Principle[59] are all similar to earlier principles. Despite the Framework's significant changes and additions to data protection standards, data deletion is mentioned only in the context of deleting inaccurate data that have been challenged by the data subject.[60] The Framework has been endorsed as the most promising set of data protection standards by several major global actors, including Google.[61]

5. Global Privacy Standard

The most recent set of proposed data protection principles is the Global Privacy Standard (GPS).[62] Adopted in 2006 by the International Data Protection Commissioners Conference, the GPS is unlike earlier sets of data protection principles that

50. *Id.* ¶ 14.
51. *Id.* ¶ 15.
52. *Id.* ¶ 18.
53. *Id.* ¶ 19.
54. *Id.* ¶ 19.
55. *Id.* ¶ 20.
56. *Id.* ¶ 21.
57. *Id.* ¶ 22.
58. *Id.* ¶ 23.
59. *Id.* ¶ 26.
60. *Id.* ¶ 23.
61. Peter Fleischer, Call for Global Privacy Standards, Google Public Policy Blog, http://googlepublicpolicy.blogspot.com/2007/09/call-for-global-privacy-standards.html (Sept. 14, 2007, 11 :03 PST).
62. ANN CAVOUKIAN, INFORMATION AND PRIVACY COMMISSIONER, CREATION OF A GLOBAL PRIVACY STANDARD 1 (2006), http://www.ipc.on.ca/images/Resources/up-gps.pdf [hereinafter GPS].

focused on economic regions. It was explicitly designed to be a definitive, global standard. The data protection commissioners sought to take account of the "strengths and weaknesses of the major codes in existence" and "harmon[ize] the principles into a single set of fair information principles."[63] Most of the principles (accountability, collection limitation, purpose, accuracy, security, openness, access, and compliance) are similar to their predecessors. However, three major changes appear.

The Consent Principle[64] explicitly deals with requiring data subject consent for the collection, use, and disclosure of data. Although consent was previously mentioned in other principles, it is now given greater significance with its own principle. Under the principle, the more sensitive the information being collected, the more specific the data subject's consent must be.[65] The Collection Limitation Principle[66] contains a data minimalization requirement, which holds that if possible, non-identifying information should be used. If identifying data must be used, then the minimal amount of data required to fulfill the purpose should be collected.

Finally, the Use, Retention, and Disclosure Limitation Principle contains the following provision: "Personal information shall be retained only as necessary to fulfill the stated purposes, and then securely destroyed."[67] This is the first general mandate to securely destroy personal information that is no longer needed by data controllers. While it is an important step toward completing data protection regulations, the provision provides a vague criterion for determining when data should be destroyed—when keeping the data is not necessary to fulfill the stated purposes of the data collection. While granting discretion to data controllers has benefits, transparency in the collection and processing of personal data and accountability of data controllers, as articulated in other principles, requires some limits on that discretion.

Data deletion has been mentioned in some sets of principles – at least obliquely, in the anonymization provisions. It is time for data deletion to become more fully developed as a complete data protection principle, like consent and choice principles, in the Framework and GPS. While procedural protections like consent and choice are important, substantive safeguards are needed to provide a floor of protection. Until the substantive protection accorded by destruction of personal data is recognized as an important part of reducing harm and respecting individual autonomy,

63. *Id.*
64. *Id.* at 3.
65. *Id.*
66. *Id.*
67. *Id.*

data protection principles will not be able to reach their full potential or serve as a solid foundation for a comprehensive, effective, global data protection regime.

II. THE NEED FOR A DATA DELETION PRINCIPLE

When there is a need for personal data, those data are collected, processed, transferred, and stored. Take the development of the credit reporting system. Lenders wanted a way to separate reliable borrowers from risky ones. By collecting the borrowing histories of potential clients, lenders could decide to whom to lend and at what interest rate. Credit reporting agencies collect information from creditors and process it (aggregating it into useful reports and producing credit scores), disclose it (sending credit reports about loan applicants to lenders), and store it for later use.

But what should happen to these data holdings when they no longer fulfill the purpose for which they were collected? For example, borrowers eventually die, and their lending histories can no longer help decide if more loans should be granted. Additionally, the failure to make a payment twenty years ago may no longer accurately reflect a borrower's current reliability. The consequences of maintaining personal information beyond its useful life are severe enough that they merit adequate preventative measures that only a complete data deletion principle can provide.

A. Consequences of Needless Data Retention: Reduction in Data Subjects' Autonomy

To respect data subjects' autonomy and human dignity, data subjects must have some measure of control over the collection, use, and disclosure of data about them. Because important decisions are made about data subjects on the basis of the data, lack of control over the data deprives individuals of control over the decisions. "Privacy involves the ability to avoid the powerlessness of having others control information that can affect whether an individual gets a job, becomes licensed to practice in a profession, or obtains a critical loan . . .. It is not merely the collection of data that is the problem—it is our complete lack of control over the ways it is used or may be used in the future."[68] Generally accepted data protection principles, like collection limitation and purpose specification, protect individual autonomy by limiting how much information is collected and requiring that the purpose be specified, but without a data deletion principle, enforcing other principles becomes much more difficult. This is particularly true because principles regulating collection and notice are generally procedural and not fully utilized by

68. SOLOVE, *supra* note 1, at 51.

individuals.[69] No one could reasonably exercise all procedural data protection rights, so substantive protections, like a requirement to destroy unneeded data, are essential to provide a baseline of data protection.

1. "Function Creep"

Data constitute a flexible resource. Financial records, for example, can be used for many reasons, such as targeting advertisements at certain economic classes, assessing taxes, or tracking money-laundering and other criminal activities. It is this malleability that makes purpose specification so crucial to data protection. Without it, institutions will inevitably be tempted to use data for new purposes, including purposes that were not specified when the data were collected.

The growing ability to aggregate, compare, and cross-check different data sets against each other creates many opportunities to use data for purposes other than the one for which the data were originally collected. For instance, the Internal Revenue Service (IRS) collects data on income for the purposes of assessing taxes.[70] The Social Security Administration (SSA) collects data on any person who applies for a social security number.[71] The Bureau of Customs and Border Protection (CBP) utilizes an Automated Targeting System to collect data on people entering and leaving the country.[72] Agencies conduct data matches between databases and combine them to create new databases.[73] Much of the value of data sold by private data aggregators, like Choicepoint and LexisNexis, derives from the combination of many other private and public data sets that, once brought together, can be used for other purposes.[74]

Function creep undermines data subjects' autonomy by depriving them, often without their knowledge, of any control over their data. Predictably, the availability of such data is also very tempting to data controllers. Why collect data all over again when you can do something useful with existing data? The history of the social security number amply illustrates the extent to which data can be used for

69. *See* Cate, *supra* note 9, 363–64, 366–67.

70. Privacy Act of 1974, as Amended; Systems of Records, 73 Fed. Reg. 13,284, 13,304 (Mar. 12, 2008).

71. Altered System of Records and New Routine Use, 63 Fed. Reg. 14,165, 14,166 (Mar. 24, 1998).

72. Notice of Privacy Act System of Records, 71 Fed. Reg. 64,543, 64,544 (Nov. 2, 2006).

73. *See, e.g.*, Dalia Naamani-Goldman, *Anti-terrorism Program Mines IRS' Records: Privacy Advocates are Concerned that Tax Data and Other Information may be Used Improperly*, L.A. TIMES, Jan. 15, 2007, at 1, *available at* 2007 WLNR 792302. For a review of government data collection and data-mining, see generally Fred. H. Cate, *Government Data Mining: The Need for a Legal Framework*, 43 HARV. C.R.-C.L. L. REV. 435 (2008).

74. Scott Canon, *Pending Merger would Combine Billions of Data Files*, KAN. CITY STAR, May 25, 2008, at A1.

purposes entirely foreign to the original collection.[75] Originally created to track employees' contributions to Social Security, the social security number is now a de facto national identifier, utilized in banks, educational institutions, and many other non-Social Security government programs. Deleting unneeded data is the surest way to avoid the temptation to deviate from the specified purposes and ensure respect for individuals' autonomy. While anonymizing unnecessary data provides some protection, if the data are not needed for their specified purpose, there is no legitimate reason to retain them. There is a risk that useful data will mistakenly be destroyed, but there is always a risk of careless records management. Since data controllers tend to default to retaining data, the risk of mistaken deletions does not outweigh the benefits of secure destruction of unneeded data.

2. Stale and Inaccurate Records

As proficient as modern data collection systems are, they are not perfect and inaccuracies are inevitable. For instance, a portion of credit bureaus' files contain important errors.[76] Every year, a number of people who are still living are recorded as deceased by the Social Security Administration (SSA), leading to enormous difficulties for both SSA and the data subjects.[77] Errors primarily creep into data sets in two ways. First, as personal data ages, they are likely to become less accurate. People frequently move, change jobs, become sick, and change their buying preferences. Second, the more data that are collected, the more likely it is that clerical errors will occur. These errors lead to mischaracterizations of individuals, thereby reducing data subjects' control over their own lives—for example, by treating them as persons that, in fact, they are not. Individual participation and access helps address this problem, but the burden for keeping records accurate and current should not fall solely on the data subject. A data deletion principle

75. *See* CHRISTIAN PARENTI, THE SOFT CAGE: SURVEILLANCE IN AMERICA FROM SLAVERY TO THE WAR OF TERROR 85–89 (2003).

76. According to one study, 79% of credit files contained errors and 25% of files had errors serious enough to result in wrongful denial of credit. *See* ALISON CASSADY & EDMUND MIERZWINSKI, NAT'L ASS'N OF STATE PUB. INTEREST RESEARCH GROUPS, MISTAKES DO HAPPEN: A LOOK AT ERRORS IN CONSUMER CREDIT REPORTS, 11–13 (2004), *available at* http://www.uspirg.org/uploads/BE/ev/BEevuvl9a3KzsATRbZMZlw/MistakesDoHappen2004.pdf.

77. OFF. OF THE INSPECTOR GEN., SOC. SEC. ADMIN., SURVIVOR BENEFITS PAID IN INSTANCES WHEN THE SOCIAL SECURITY ADMINISTRATION REMOVED THE DEATH ENTRY FROM A PRIMARY WAGE EARNER' S RECORD, Audit Rep. A-06–06–26020 (Sept. 2006) 1–2, *available at* http://www.ssa.gov/oig/ADOBEPDF/A-06–06–26020.pdf; Alex Johnson & Nancy Amons, *Government Records Incorrectly Kill off Thousands, and There's No Easy Fix*, MSNBC, Feb. 29, 2008, *availabe at* http://www.msnbc.msn.com/id/23378093/.

helps eliminate stale and inaccurate records by destroying the very records that are most likely to be irrelevant, erroneous, or out-of-date.

B. Consequences of Needless Data Retention: Increased Risk of Harm to Individual Data Subjects

In addition to undermining individuals' dignity and reducing their control over important life decisions made about them, retaining unneeded data increases the risk of the data being misused to the detriment of the data subjects. As data holdings have become more extensive, they are more valuable to criminals who are capable of using the data to commit fraud.

1. Data Theft, Loss, and Fraud

Personal information has become a valuable commodity for people who seek to defraud banks, insurance companies, and other corporations. Another individual's data can be used to fraudulently obtain credit, steal money, avoid the consequences of a criminal record, cast multiple votes in an election, or illegally gain employment.[78]

While relatively old data may not be useful for some illegal purposes (a ten-year-old credit card number is unlikely to be valid), some, like a social security number, can remain very useful. Even the records of deceased individuals can be used; sometimes, they are more valuable because there is no living victim to complain about the fraud. The longer a data controller retains data, the more likely that data will be lost or misused. If old data are not destroyed, they accumulate, increasing the opportunity for loss because there are more data to catalog and secure, and it becomes more difficult to securely transport and store the data. Old data, especially data that are not being kept for a specific purpose, are less likely to be used regularly, and thus, security measures may be more lax and breaches more difficult to promptly detect. When data breaches do occur, they can cause more damage because they affect more data subjects whose data, if regular data destruction were practiced, would have been deleted.

In recent years, numerous data breaches have been reported, some of which contained rather old data and some of which involved data controllers disposing of records in a way that left the data exposed to misappropriation.[79] Even if data are merely

78. *See, e.g.*, Alexia Elejalde-Ruiz, *Identity Crisis: As Society Goes More Hi-tech, so do the Thieves of Personal Information*, CHI. TRIB., Aug. 18, 2008, at 6, *available at* 2008 WLNR 15616650.

79. For extensive examples of data breaches, many of which involve old or improperly destroyed data, see The Breach Blog, http://breachblog.com/ (last visited Aug. 18, 2008) and Privacy News, http://www.pogowasright.org/ (last visited Aug. 18, 2008).

misplaced or lost, data subjects then have to live with the anxiety of potentially being harmed through use of the lost data. Under a data deletion principle, some of that unneeded data would be destroyed, and therefore could not be lost or misused.

2. Harmful Decisions Made on The Basis of Stale and Inaccurate Data

As discussed above, the longer information is retained, the more difficult it is to keep the data relevant and accurate. Since personal information is used to make extremely important decisions about individuals, inaccurate information can have catastrophic consequences for data subjects. Inaccurate medical records can lead to dangerous treatment decisions; incorrect credit records can result in adverse loan decisions; and erroneous criminal records can result in unwarranted job terminations.[80] Personal data are useful only to the extent that they are accurate and current. Thus, a data deletion principle is important not only for reducing the risk of data subjects being harmed by misuse of their data, but also for increasing the probability that the data will be suitable for their beneficial purposes.

III. A PROPOSED DATA DELETION PRINCIPLE

Global data protection principles, once universally accepted, will help harmonize national data protection laws, thereby facilitating efficient, transnational transfers of personal data that serve important social and economic functions. Unfortunately, data protection principles have thus far given short shrift to data deletion as a crucial component of an effective global data protection regime. This omission has left a gap in data protection norms, reducing data controllers' respect for data subjects' autonomy and exposing data subjects to potential abuse of their data. The Data Deletion Principle (hereinafter the Principle) proposed below aims to fill that gap, and its inclusion will make a global data protection regime more robust by regulating the end of information's life cycle.

A. Text of the Data Deletion Principle

DATA DELETION—All personal information, regardless of format, data controller, or location, should be securely and verifiably destroyed within:

80. SOLOVE, *supra* note 1, at 46–47.

a. the time specified in a publicly available data destruction schedule that has been approved by a legally authorized data protection authority, provided that the time period is necessary for the specified purpose of the data and does not exceed ten years after creation of the record; or
b. one year after creation of the record, if the record is not subject to a data destruction schedule; unless,
c. the data protection authority approves a longer retention period, including permanent retention.

The data controller shall have the burden of showing that the longer retention period is justified by serving the public good or the interests of the data subjects. Data shall be destroyed in accordance with the data destruction schedule approved in the nation in which the data are collected. Failure to destroy personal information within the applicable period should render the data controller liable to any data subjects who are harmed by misuse of information that should have been deleted. National data protection authorities should ensure that data destruction schedules are consistent across industries and jurisdictions. Additionally, the data protection authorities should enforce compliance through appropriate legal mechanisms, excluding data controllers with fewer than twenty-five employees and holding personal data on fewer than five hundred individuals.

B. Implementation of the Data Deletion Principle

Compliance with the Principle will impose administrative burdens on data controllers, but it is consistent with good record management practices and existing legal data destruction obligations.[81] Despite additional cost, securely destroying unneeded data is simply a part of responsible data management. As such, data deletion should be considered as important as other data management practices.

1. Application of the Data Deletion Principle

Since personal records are held by both government and private sector data controllers, the Principle must apply equally to both public and private actors. However, due to the amount of resources required, very small data controllers, those that em-

81. *See* ANDREW B. SERWIN, INFORMATION SECURITY AND PRIVACY: A PRACTICAL GUIDE TO FEDERAL, STATE AND INTERNATIONAL LAW §19:9 (2007); PETER P. SWIRE & SOL BERMANN, INFORMATION PRIVACY: OFFICIAL REFERENCE FOR THE CERTIFIED INFORMATION PRIVACY PROFESSIONAL [CIPP] 173–74 (2007).

ploy fewer than twenty-five people and control data on fewer than five hundred individuals, will not be subject to enforcement actions by national data protection authorities. While these small data controllers will not be penalized by data protection authorities for neglecting to file data destruction schedules or to destroy data, they are still liable to individuals who are harmed by abuse of data that are not securely destroyed within the applicable time limits. Imposition of liability provides some incentive to comply with the Principle, and may also induce insurance companies to encourage or require compliance as a condition of issuing liability insurance.

2. Data Destruction Schedules

Transparency, accountability, and individual choice are served by requiring data controllers to catalog and publicly declare their personal data systems and to set definite retention periods before data will be securely destroyed. Data subjects will be able to know how long data about them are retained, thus increasing their awareness of the bureaucratic processes that handle the data. Like current privacy policies, it is unlikely that data destruction schedules will be read by many data subjects.[82] However, the most basic level of individual access is making information about data collection, processing, and disposal publicly available. Public destruction schedules will also facilitate oversight by private actors and data protection authorities.

All data destruction schedules should be filed with, and approved by, authorized data protection authorities. While most nations have government agencies that serve this function, some may opt to authorize a private or quasi-governmental entity to approve data destruction schedules. It is important that the authority vigorously review the schedules and ensure consistency across industries. With strong enforcement, industry norms that roughly standardize data retention periods would likely be developed, and the authority would then enforce those norms by rejecting outlier schedules. For example, if banks generally retained certain transaction data for five years, the authority would require special justification before approving a schedule that retains the same data for fifteen years. Just as telemarketers pay fees to fund federal and state do-not-call registries, review of data destruction schedules could be funded by the data controllers.

3. Data Retention Periods

The Principle sets a short retention period for transitory data and thus relieves data controllers from having to catalog it. Any data needed for more than a year will be cataloged in a schedule. However, data controllers cannot be permitted simply to

82. Cate, *supra* note 9, at 360–61.

impose whatever time limit they wish. A data destruction schedule that merely declares that all data will be kept indefinitely is not particularly helpful in terms of preventing the harms caused by retention of unneeded, stale data. Hence, the data protection authority must review the retention periods within the submitted data destruction schedules and ensure that they are necessary to achieve the purposes for which the data are collected. Admittedly, some data (though probably a relatively small portion of all data collected) will have to be kept for a long time, perhaps even permanently. Retention periods over ten years must be justified by the data subjects' interests or the public interest. Because the retention period is tied to the specified purpose of the data collection, the risk of function creep is greatly reduced. Data will be destroyed before new, unrelated uses for the data appear.

4. Secure Destruction

When the retention period for data ends, it is critical that the data be securely and verifiably destroyed. Otherwise, the data will be exposed to misuse. Verification of data destruction must be maintained to permit oversight by the data protection authority and to protect the data controller from liability to data subjects. Mechanisms for secure data destruction, such as shredders and degassers, already exist to comply with the few data destruction mandates that are currently utilized.

Some personal data can be useful in an anonymized or aggregate form. Such data, assuming they could not be used to identify individuals, would no longer be personal information and would not be regulated by data protection principles. As long as the identifying components of the data are securely deleted, aggregate data derived from personal data can be used and retained for any reason.

5. Consistency Across Industries and Jurisdictions

A major challenge for the implementation of harmonized national data deletion mandates is maintaining relatively consistent data destruction schedules across industries and jurisdictions. Unless a supranational data protection authority is created (a project even more ambitious than establishing universal data protection principles), there will inevitably be some variation among the data protection authorities' handling of data destruction schedules. This is a difficulty that is inherent in a system that adopts universal norms while granting discretion in implementation to each state, and must be accepted at this point in the development of a global data protection regime.

Variation can be minimized between data controllers in each nation by the data protection authorities. Roughly similar retention periods should be required

for similar systems of personal information collected for like purposes. Multinational corporations should apply similar retention periods within all their subsidiaries, regardless of jurisdiction. This will foster consistency across jurisdictions.

This proposal assumes a level of cooperation between national data protection authorities. A conflict could arise if industry standards in one jurisdiction are too different from legal mandates in another. For instance, search engines in the United States have begun to delete identifying data in search logs after eighteen months.[83] At the same time, some European nations are considering data retention laws that require storage of internet usage data for a longer period.[84] While these two examples do not directly conflict, they indicate diverging balances between privacy and data retention. Conflicts between these jurisdictions may put data controllers in an untenable position of being unable to have approved schedules in both countries.

This is an unavoidable consequence of divergent approaches to data protection and can be solved only by negotiation and compromise among the national data protection authorities. It is impossible to have a harmonized, global data deletion mandate while permitting wide disparities among jurisdictions' retention standards. Adoption of a data deletion principle does not end the global dialogue on data protection; deliberation must continue to establish more specific global retention standards.

6. Enforcement

Enforcement of the Principle can be accomplished by national data protection authorities and the individual data subjects who have been harmed by the misuse of their personal information. Data protection authorities can periodically audit data controllers' verification records and determine if the data destruction schedules are being followed. Fines, regulatory orders, and other administrative measures can help ensure compliance. Data protection authorities will likely focus their resources on large and systemic violations, while an individual right of action for data subjects harmed by misuse of improperly handled data will hold controllers responsible for smaller violations that lead to actual harm. Large data breaches, like the TJ Maxx breach,[85] if involving data that were not destroyed in accordance with data deletion laws, would subject the controller to heavy liability

83. Miguel Helft, *Google Adds a Safeguard on Privacy for Searchers*, N.Y. TIMES, Mar. 15, 2007, at C4.

84. Matthew Sparkes, *Government Proposes Email and Internet Tracking*, PCPRO, Aug. 13, 2008, http://www.pcpro.co.uk/news/218052/government-proposes-internet-tracking.html.

85. Eric Dash, *Data Breach Could Affect Millions of TJX Shoppers*, N.Y. TIMES, Jan. 19, 2007, at C9.

to harmed individuals, giving individuals redress and providing a strong incentive to controllers to destroy data in accordance with applicable retention periods.

The enforcement mechanisms directly relate to the purposes of the Principle: protecting individual autonomy and avoiding harm. Data protection authorities' enforcement activities will encourage data destruction norms, protect individual autonomy, and increase public confidence in data collection, processing, and global transfer. Furthermore, individual lawsuits will help discourage careless record disposal and provide redress for harms caused by abuse facilitated by insecurely discarded data.

CONCLUSION

With digital capture, processing, transfer, and storage of personal information becoming less expensive, it is easy for data controllers to adopt a "packrat" attitude toward data retention. Why dispose of information when it is cheap to keep and you never know when it might come in handy? However, this attitude undermines generally accepted data protection norms and deprives data subjects of knowledge and control over important decisions made on the basis of their data. It also increases the chances that the data will be used in ways that deviate from the data's original purpose and that it will be stolen, lost, or misused to harm both data subjects and controllers.

The tendency to accumulate and keep personal information indefinitely must be resisted if global data protection principles are to be as effective as possible. A data deletion principle is needed that encourages public disclosure of data retention periods, that is related to the data's purposes, and that is generally consistent across borders between nations and industries. Unneeded data should be securely and verifiably destroyed; failure to do so should subject most controllers to administrative penalties and liability to harmed individuals.

Global transfers of personal data are important and should be encouraged. Data protection standards that regulate the entire data life cycle, from collection to deletion, will increase data subjects' confidence in data controllers and lower legal barriers to moving data across national borders. Collection of personal data is a global phenomenon, and a global data protection regime that includes data deletion requirements is best suited to harness the advantages of global data while minimizing the risks of harm.