Panel Discussion on Key Privacy Issues

Omer Tene

Dec 12, 2019

CODE

Balancing Privacy with Health Data Access

Roundtable Report

September 2019

GLOBAL PULSE

ABOUT
PROJECTS
LABS
NEWS
CHALLENGES
PRIVACY
PARTNERSHIPS
RESOURCES
CONTACT
HOME

BLOG

UNPACKING THE ISSUE OF MISSED USE AND MISUSE OF DATA

Robert Kirkpatrick, Director, UN Global Pulse *Mar 18, 2019*

Just because data misuse is at the forefront of recent conversations, we shouldn't ignore the harms associated with missed use. Lost opportunities to use big data to achieve the Sustainable Development Goals (SDGs) are probably to blame for at least as much harm as leaks and privacy breaches.

iapp

BUILDING ETHICS INTO PRIVACY FRAMEWORKS FOR BIG DATA AND AI

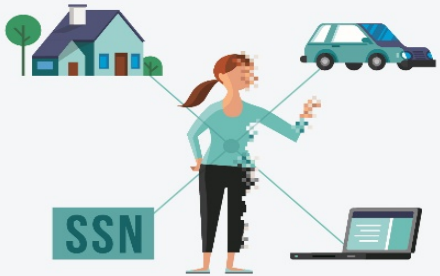
A VISUAL GUIDE TO PRACTICAL DATA DE-IDENTIFICATION

Produced by FUTURE OF PRIVACY FORUM FPF.ORG

In collaboration with EY

What do scientists, regulators and lawyers mean when they talk about de-identification? How does anonymous data differ from pseudonymous or de-identified information? Data identifiability is not binary. Data lies on a spectrum with multiple shades of identifiability.

This is a primer on how to distinguish different categories of data.



DEGREES OF IDENTIFIABILITY

Information containing direct and indirect identifiers.

PSEUDONYMOUS DATA

Information from which direct identifiers have been eliminated or transformed, but indirect identifiers remain intact.

DE-IDENTIFIED DATA

Direct and known indirect identifiers have been removed or manipulated to break the linkage to real world identities.

ANONYMOUS DATA

Direct and indirect identifiers have been removed or manipulated together with mathematical and technical guarantees to prevent re-identification.

	EXPLICITLY PERSONAL	POTENTIALLY IDENTIFIABLE	NOT READILY IDENTIFIABLE	KEY CODED	PSEUDONYMOUS	PROTECTED PSEUDONYMOUS	DE-IDENTIFIED	PROTECTED DE-IDENTIFIED	ANONYMOUS	AGGREGATED ANONYMOUS
DIRECT IDENTIFIERS Data that identifies a person without additional information or by linking to information in the public domain (e.g., name, SSN)	INTACT	PARTIALLY MASKED	PARTIALLY MASKED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED
INDIRECT IDENTIFIERS Data that identifies an individual indirectly. Helps connect pieces of information until an individual can be singled out (e.g., DOB, gender)	INTACT	INTACT	INTACT	INTACT	INTACT	INTACT	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED
SAFEGUARDS and CONTROLS Technical, organizational and legal controls preventing employees, researchers or other third parties from re-identifying individuals	NOT RELEVANT due to nature of data	LIMITED or NONE IN PLACE	CONTROLS IN PLACE	CONTROLS IN PLACE	LIMITED or NONE IN PLACE	CONTROLS IN PLACE	LIMITED or NONE IN PLACE	CONTROLS IN PLACE	NOT RELEVANT due to nature of data	NOT RELEVANT due to high degree of data aggregation
SELECTED EXAMPLES	Name, address, phone number, SSN, government-issued ID (e.g., Jane Smith, 123 Main Street, 555-555-5555)	Unique device ID, license plate, medical record number, cookie, IP address (e.g., MAC address 68:A8:6D:35:65:03)	Same as Potentially Identifiable except data are also protected by safeguards and controls (e.g., hashed MAC addresses & legal representations)	Clinical or research datasets where only curator retains key (e.g., Jane Smith, diabetes, HgB 15.1 g/dl = Csrk123)	Unique, artificial pseudonyms replace direct identifiers (e.g., HIPAA Limited Datasets, John Doe = 5L7T LX619Z) (unique sequence not used anywhere else)	Same as Pseudonymous, except data are also protected by safeguards and controls	Data are suppressed, generalized, perturbed, swapped, etc. (e.g., GPA: 3.2 = 3.0-3.5, gender: female = gender: male)	Same as De-Identified, except data are also protected by safeguards and controls	For example, noise is calibrated to a data set to hide whether an individual is present or not (differential privacy)	Very highly aggregated data (e.g., statistical data, census data, or population data that 52.6% of Washington, DC residents are women)

DOI:10.1145/3287287

Article development led by acmqueue
queue.acm.org

These attacks on statistical databases are no longer a theoretical danger.

BY SIMSON GARFINKEL, JOHN M. ABOWD, AND CHRISTIAN MARTINDALE

Understanding Database Reconstruction Attacks on Public Data

IN 2020, THE U.S. Census Bureau will conduct the Constitutionally mandated decennial Census of Population and Housing. Because a census involves collecting large amounts of private data under the promise of confidentiality, traditionally statistics are published only at high levels of aggregation. Published statistical tables are vulnerable to *database reconstruction attacks* (DRAs), in which the underlying microdata is recovered merely by finding a set of microdata that is consistent with the published statistical tabulations. A DRA can be performed by using the tables to create a set of mathematical constraints and then solving the resulting set of simultaneous equations. This article shows how such an attack can be addressed by adding noise to the published tabulations, so the reconstruction no longer results in the original data. This has implications for the 2020 census.

The goal of the census is to count every person once, and only once, and in the correct place. The results are used to fulfill the Constitutional requirement to apportion the seats in the U.S. House of Representatives among the states according to their respective numbers.

In addition to this primary purpose of the decennial census, the U.S. Congress has mandated many other uses for the data. For example, the U.S. Department of Justice uses block-by-block counts by race for enforcing the Voting Rights Act. More generally, the results of the decennial census, combined with other data, are used to help distribute more than $675 billion in federal funds to states and local organizations.

Beyond collecting and distributing data on U.S. citizens, the Census Bureau is also charged with protecting the privacy and confidentiality of survey responses. All census publications must uphold the confidentiality standard specified by Title 13, Section 9 of the U.S. Code, which states that Census Bureau publications are prohibited from identifying "the data furnished by any particular establishment or individual." This section prohibits the Census Bureau from publishing respondents' names, addresses, or any other information that might identify a specific person or establishment.

Upholding this confidentiality requirement frequently poses a challenge, because many statistics can inadvertently provide information in a way that can be attributed to a particular entity. For example, if a statistical agency *accurately* reports there are two persons living on a block and the average age of the block's residents is 35, that would constitute an improper disclosure of personal information, because one of the residents could look up the data, subtract their contribution, and infer the age of the other.

Subsec. (a). Pub. L. 94–521 substituted provision that the Secretary may furnish to any respondent, or the successor or authorized agent of such respondent, transcripts or copies of reports containing information furnished in connection with the surveys and census, upon payment of the necessary costs, for provision that authorized the Secretary, in his discretion, to furnish the Governors of States and Territories, courts of record, and individuals, data for genealogical and other proper purposes, from the population, agriculture, and housing schedules prepared under the authority of subchapter II of chapter 5 of this title, upon payment of the necessary costs, plus one dollar for supplying a certificate.

Subsec. (b). Pub. L. 94–521 inserted provision subjecting the Secretary to the limitations contained in sections 6(c) and 9 of this title, when furnishing statistical materials under this section, substituted ''copies of tabulations and other statistical materials'' for ''transcripts or copies of tables and other census records'', inserted provision that materials furnished under this section may not disclose information reported by, or on behalf of, a particular respondent, and substituted a provision enumerating the public and private establishments and individuals, on behalf of whom, special statistical compilations may be conducted for provision that such compilations may be conducted on behalf of State or local officials, private concerns, or individuals.

Subsec. (c). Pub. L. 94–521 struck out ''the authority of'' after ''furnished under'', substituted ''any respondent or other person'' for ''the persons'', and inserted ''except in the prosecution of alleged violations of this title'' after ''relates,''.

1957—Subsec. (b). Pub. L. 85–207, §4(a), inserted sentence at end respecting engagement in joint statistical projects.

Subsec. (d). Pub. L. 85–207, §4(b), required the deposit in a separate account of moneys received in payment for work or services, previously credited to an appropriation for collecting statistics, and permitted certain uses of such account.

EFFECTIVE DATE OF 1976 AMENDMENT

Amendment by Pub. L. 94–521 effective Oct. 17, 1976, see section 17 of Pub. L. 94–521, set out as a note under section 1 of this title.

§ 9. Information as confidential; exception

(a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison, may, except as provided in section 8 or 16 or chapter 10 of this title or section 210 of the Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations Act, 1998 or section 2(f) of the Census of Agriculture Act of 1997—

(1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or

(2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or

(3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports.

No department, bureau, agency, officer, or employee of the Government, except the Secretary in carrying out the purposes of this title, shall require, for any reason, copies of census reports which have been retained by any such establishment or individual. Copies of census reports which have been so retained shall be immune from legal process, and shall not, without the consent of the individual or establishment concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding.

(b) The provisions of subsection (a) of this section relating to the confidential treatment of data for particular individuals and establishments, shall not apply to the censuses of governments provided for by subchapter III of chapter 5 of this title, nor to interim current data provided for by subchapter IV of chapter 5 of this title as to the subjects covered by censuses of governments, with respect to any information obtained therefor that is compiled from, or customarily provided in, public records.

(Aug. 31, 1954, ch. 1158, 68 Stat. 1013; Pub. L. 87–813, Oct. 15, 1962, 76 Stat. 922; Pub. L. 101–533, §5(b)(2), Nov. 7, 1990, 104 Stat. 2348; Pub. L. 103–430, §2(b), Oct. 31, 1994, 108 Stat. 4394; Pub. L. 105–113, §4(a)(1), Nov. 21, 1997, 111 Stat. 2276; Pub. L. 105–119, title II, §210(k), Nov. 26, 1997, 111 Stat. 2487.)

HISTORICAL AND REVISION NOTES

Based on title 13, U.S.C., 1952 ed., §§73, 83, 122, 208, 211, 252, and section 1442 of title 42, U.S.C., 1952 ed., The Public Health and Welfare (Aug. 7, 1916, ch. 274, §3, 39 Stat. 437; Apr. 2, 1924, ch. 80, §3, 43 Stat. 31; June 18, 1929, ch. 28, §§8, 11, 21, 46 Stat. 23, 25, 26; July 25, 1947, ch. 331, 61 Stat. 457; June 19, 1948, ch. 502, §2, 62 Stat. 479; July 15, 1949, ch. 338, title VI, §607, 63 Stat. 441; Sept. 7, 1950, ch. 910, §2, 64 Stat. 784).

Section consolidates parts of sections 73 and 83 of title 13, U.S.C., 1952 ed., part of section 208 of such title, section 211 of such title, that part of section 122 of such title which made such sections 208 and 211 applicable to the quinquennial censuses of manufacturers, the mineral industries, and other businesses (see subchapter I of chapter 5 of this revised title), that part of section 252 of such title which made such sections 208 and 211 applicable to the quinquennial censuses of governments (see subchapter III of chapter 5 of this revised title), the second proviso in such section 252, and that part of subsection (b) of section 1442 of title 42, U.S.C., 1952 ed., which made such sections 208 and 211 applicable to the decennial censuses of housing (see subchapter II of chapter 5 of this revised title).

Words ''except as provided in section 8 of this title'' were inserted in opening phrase of subsection (a) for the purpose of clarity.

References to the Secretary, the Department of Commerce and bureaus and agencies thereof, and to other officers and employees of such Department, bureaus or agencies, were substituted for references to the Director of the Census, the ''Census Office'', and the enumeration (in section 208 of title 13, U.S.C., 1952 ed.) of certain types of employees, for the purpose of completeness, and to conform with 1950 Reorganization Plan No. 5, §§1, 2, eff. May 24, 1950, 15 F.R. 3174, 64 Stat. 1263. See Revision Note to section 4 of this title.

The penal provisions of sections 73, 83, and 208 of title 13, U.S.C., 1952 ed., prescribing penalties for wrongful disclosure of information, are set out in section 214 of this title.

Changes were made in phraseology.

For remainder of sections 122, 208, and 252 of title 13, U.S.C., 1952 ed., and of section 1442 of title 42, U.S.C., 1952 ed. (which section has been transferred in its entirety to this revised title), see Distribution Table.

REFERENCES IN TEXT

Section 210 of the Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations Act, 1998, referred to in subsec. (a), is section 210 of Pub. L. 105–119, title II, Nov. 26, 1997, 111 Stat.

HIPAA

- **§ 164.514 Other requirements relating to uses and disclosures of protected health information.**

- (a) **Standard: De-identification of protected health information.** Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information (IIHI).

- **(b) Implementation Specifications: Requirements for de-identification of PHI.** A covered entity may determine that health information is not IIHI using one of these approaches
- (1) Expert determination
- (2) Safe harbor

GDPR

- **Article 4(1) ‘personal data’** means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier ... or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.
- **Recital 26** To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly.

CCPA

- **§ 1798.140(o)(1) “Personal information”** means information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household.
- **(3) “Personal information”** does not include consumer information that is deidentified or aggregate consumer information.

Thank You!