Panel Discussion on Key Privacy Issues

Omer Tene
Dec 12, 2019
Balancing Privacy with Health Data Access

Roundtable Report

September 2019
UNPACKING THE ISSUE OF MISSED USE AND MISUSE OF DATA

Robert Kirpatrick, 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.
BUILDING ETHICS INTO PRIVACY FRAMEWORKS FOR BIG DATA AND AI
A VISUAL GUIDE TO PRACTICAL DATA DE-IDENTIFICATION

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.

- **Explicitly Personal**
- **Potentially Identifiable**
- **Not Readily Identifiable**

**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**
- **Not Relevant due to nature of data**

**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**
- **Not Relevant due to nature of data**

**SAFEGUARDS and CONTROLS**
Technical, organizational, and legal controls preventing employees, researchers or other third parties from re-identifying individuals.

- **Controls in Place**
- **Limited or None in Place**

**SELECTED EXAMPLES**
- Name, address, phone number, SSN, government-issued ID (e.g., driver's license)
- Unique device ID, license plate, medical record number, credit score, IP address
- Clinical or research datasets where only curator retains key
- Unique, artificial pseudonyms replace direct identifiers (e.g., HIPAA Limited Data Set)
- Data are suppressed, generalized, perturbed, or re-mapped (e.g., GDN, K-3, K-5, gender: female -> gender: male)

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

- **Elminated or Transformed**
- **Controls in Place**
- **Limited or None in Place**

**DE-IDENTIFIED DATA**
Direct and known indirect identifiers have been removed or manipulated to break the linkage to real world identities.

- **Elminated or Transformed**
- **Controls in Place**
- **Limited or None in Place**

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

- **Elminated or Transformed**
- **Controls in Place**
- **Limited or None in Place**

**AGGREGATED ANONYMOUS**
Not relevant due to high degree of data aggregation.

- **Controls in Place**
- **Not Relevant due to nature of data**
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.
§ 9. Information as confidential; exception

(a) Neither the Secretary, nor any officer or employee of the Department of Commerce or bureau or agency thereof, or local government or political subdivision, may, except as provided in section 8 or 10 or chapter 10 of this title or section 210 of the Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations Act, 1999 or section 217 of the Census of Agriculture Act of 1997—

(1) use the information furnished by the Secretary to the Secretary for any purpose other than the statistical purposes for which it is supplied;

(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 subsection (b) of this section.

(c) 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 subsection (b) of this section.

(d) 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 subsection (b) of this section.

(e) 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 subsection (b) of this section.

(f) 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 subsection (b) of this section.

(g) 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 subsection (b) of this section.

(h) 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 subsection (b) of this section.

(i) 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 subsection (b) of this section.

(j) 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 subsection (b) of this section.

(k) 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 subsection (b) of this section.

(l) 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 subsection (b) of this section.

(m) 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 subsection (b) of this section.

(n) 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 subsection (b) of this section.

(o) 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 subsection (b) of this section.

(p) 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 subsection (b) of this section.

(q) 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 subsection (b) of this section.

(r) 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 subsection (b) of this section.

(s) 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 subsection (b) of this section.

(t) 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 subsection (b) of this section.

(u) 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 subsection (b) of this section.

(v) 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 subsection (b) of this section.

(w) 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 subsection (b) of this section.

(x) 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 subsection (b) of this section.

(y) 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 subsection (b) of this section.

(z) 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 subsection (b) of this section.
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!