CENTER FOR DEMOCRACY & TECHNOLOGY

CDT'S GUIDE TO DEFINING TECHNICAL TERMS IN STATE PRIVACY LEGISLATION

Introduction

State legislators and regulators face unprecedented privacy and security policy issues related to new technologies. These issues include everything from updating 911 systems to deploying new technologies for use in law enforcement. As lawmakers wrestle with adapting, writing, and interpreting laws around modern technology, they confront the challenge of accurately describing rapidly evolving technology while defining appropriate accountability measures and effective privacy protections.

Unfortunately, the dynamic and evolving nature of technology make it a difficult area in which to legislate. In a pinch, legislators and their staff may turn to model/uniform bills, lobbyists, advocates, and existing federal definitions for guidance on how to define and explain data uses, data processes, and other technologies. But these sources may be highly context dependent, outdated, and/or a given definition may not necessarily align with a particular legislator's intent or objectives.

Accurate definitions of key technologies, processes, or subject areas, then, are critical to enabling state lawmakers to express their legislative intent, correctly scope implementation of a law, and effectively protect personal privacy. Such definitions should be technically-sound, durable, accurate, and provide options for achieving the intended results of the bill's author.

How to Use this Document

This document provides background information and different approaches for defining and implementing into legislation key technology and privacy-related terms. In addition to offering potential definitions for each term, we discuss the foundational features of the relevant technology, process, or concept in order to help legislators incorporate the terms in a manner that meets their goals and can be implemented in various legislative contexts.

Universal definitions in technology or legislation are generally not feasible, as the scope or specificity of a term may need to change depending on the content and purpose of the bill. For example, a bill that is renewed regularly may use a specific definition of encryption (with the knowledge that it will be updated as the technology evolves), while a bill that is meant to be enforced indefinitely might seek a less specific definition. To address this problem, this document considers the features of the relevant technologies or processes that legislation should take into account when defining key terms, while also

providing contextual background information such as common uses of relevant technologies, technical definitions, and discussions of outcomes-based (e.g., "data will be secure") versus process-based uses (e.g., "data should be made secure using encryption).

Technology and Privacy Definitions for State Legislators

ENCRYPTION

Encryption overview

Encryption is a technical process that prevents access to information by unauthorized parties**.** Any time data is transmitted or stored electronically, its security and integrity are at risk. While no tool can offer complete security, encryption is indispensable as a basic security measure to protect information networks from malicious hacking, breach and disclosure of personal information, identity theft, harassment, and other crimes.

Encryption is a technology built upon a mathematical process. Any definition should include reference to an algorithm or cryptographic process that creates a low probability of decrypting, or exposing, the information. Some definitions of encryption focus on specific technical standards, such as the Advanced Encryption Standard (AES), which is one of several cryptographic algorithms approved by the National Institute of Standards and Technology (NIST). Other definitions are outcome-driven, meaning that the information is practically inaccessible by an unauthorized party, even if it's not technically encrypted. Your needs will depend on which of these approaches makes sense for the policy you are attempting to implement.

Defining encryption

From Nevada's NRS 603A.215):

> (b) "Encryption" means the protection of data in electronic or optical form, in storage or in transit, using:
>
> > (1) An encryption technology that has been adopted by an established standards setting body, including, but not limited to, the Federal Information Processing Standards issued by the National Institute of Standards and Technology, which renders such data indecipherable in the absence of associated cryptographic keys necessary to enable decryption of such data;

Definitions of encryption should also include a definition of an an encryption key and an adequacy standard:

1. Encryption key (From California AB 2828) "encryption key" is defined as "the confidential key or process designed to render the data useable, readable, and decipherable."
2. Adequacy standard (From Maryland Senate Bill 525)
   (c) "Encrypted" means the protection of data in electronic or optical form, in storage or in transit, using an encryption technology that:

(1) Has been adopted by an established standards-setting body of the federal government, including the Federal Information Processing Standards issued by the National Institute of Standards and Technology; and
(2) Renders the data indecipherable without an associated cryptographic key necessary to enable decryption of the data.

DELETION

Deletion overview
There are different ways both to think about and categorize how information is destroyed. What it means to "delete" data, for example, runs the spectrum from simply removing the ability to link to data on a drive to the physical destruction of the drive itself.

The National Institute of Standards and Technology (NIST) defines data destruction as a process that renders paper records unreadable and, more importantly, digital data irretrievable. It results in electronic information that is forgotten, erased, deleted, completely or reliably removed, purged, sanitized, revoked, or destroyed. Legal definitions can describe both the final result that information be irretrievable as well as the procedural elements for eliminating information (e.g. overwriting data on the drive).

Deletion is not binary — data isn't either deleted or not. Rather than presenting a black-and-white view of deletion, legislation can recognize that deletion is often a matter of the *probability of recovery.* For example, de-indexing files on a drive is insufficient as a standard for data destruction because there is a high probability of recovery with little expertise or equipment.

NIST has written about two methods of data recovery: keyboard attacks and laboratory attacks. Less sensitive data should be eliminated such that a recovery attempt using readily available software recovery methods is unlikely to succeed. More sensitive data, however, should be destroyed such that even advanced laboratory attacks using specialized equipment are unlikely to recover the data. Legislation can reflect this context-specific standard for the likelihood of recovery depending on the sensitivity of the data in question.

NIST description of storage medium destruction:

> There are many different types, techniques, and procedures for media destruction. If destruction is decided on because of the high security categorization of the information, then after the destruction, the media should be able to withstand a laboratory attack.

Disintegration, Pulverization, Melting, and Incineration.

> These sanitization methods are designed to completely destroy the media. They are typically carried out at an outsourced metal destruction or licensed incineration facility with the specific capabilities to perform these activities effectively, securely, and safely.

Defining deletion

From Illinois' general Personal Information Protection Act:

> Electronic media and other non-paper media containing personal information may be destroyed or erased so that personal information cannot practicably be read or reconstructed.

A more detailed, process-based definition can be found in Illinois Data Security on State Computers Act

> Sec. 20. Establishment and implementation.
>
> The Data Security on State Computers Act is established to protect sensitive data stored on State-owned electronic data processing equipment to be (i) disposed of by sale, donation, or transfer or (ii) relinquished to a successor executive administration. This Act shall be administered by the Department or an authorized agency. The governing board of each public university in this State must implement and administer the provisions of this Act with respect to State-owned electronic data processing equipment utilized by the university. The Department or an authorized agency shall implement a policy to mandate that all hard drives of surplus electronic data processing equipment be erased, wiped, sanitized, or destroyed in a manner that prevents retrieval of sensitive data and software before being sold, donated, or transferred by (i) overwriting the previously stored data on a drive or a disk at least 3 times or physically destroying the hard drive and (ii) certifying in writing that the overwriting process has been completed by providing the following information: (1) the serial number of the computer or other surplus electronic data processing equipment; (2) the name of the overwriting software or physical destruction process used; and (3) the name, date, and signature of the person performing the overwriting or destruction process. The head of each State agency shall establish a system for the protection and preservation of State data on State-owned electronic data processing equipment necessary for the continuity of government functions upon it being relinquished to a successor executive administration.

PERSONALLY IDENTIFIABLE INFORMATION

PII overview

Information that is or can be connected to an identifiable individual is often referred to as "personally identifiable information" or PII. While most privacy legislation applies to PII or an equivalent term, not every definition of PII is the same. Many definitions include long lists of information that often vary across different laws and regulations. Some state data breach notification statutes are only triggered when a person's name is disclosed in combination with other pieces of data, while other laws protect any personal information that can be tied to an individual. Moreover, some legislation only protects "sensitive" PII.

Defining and scoping PII

CENTER FOR DEMOCRACY & TECHNOLOGY

There are two primary ways of conceptualizing PII. The first way is to define PII in terms of its "linkability" to an individual—that is, how easy it is to connect the information to an identifiable individual. This is an important component of defining PII because removing names, addresses, and other traditional identifying information from personal data may not prevent re-identification; for example, it is often possible to re-identify data by comparing multiple publicly available databases.

From the (repealed) Federal Communications Commission's broadband privacy order:

> "We define personally identifiable information, or PII, as any information that is linked or reasonably linkable to an individual or device."

Personally Identifiable Data (from California AG report)

> "Are any data linked to a person or persistently linked to a mobile device: data that can identify a person via personal information or a device via a unique identifier. Included are user-entered data, as well as automatically collected data."

Personally Identifiable Information II. (from GAO report Alternatives Exist for Enhancing Protection of Personally Identifiable Information)

> "...any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, Social Security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information."

Another way to conceptualize PII is to create a list of types of data that comprise PII. State legislation runs the gambit from narrow to broadly scoped lists.

Personally Identifiable Information I. (from Cal. Bus. & Prof. Code §22577(a))

> (1) A first and last name.
> (2) A home or other physical address, including street name and name of a city or town.
> (3) An e-mail address.
> (4) A telephone number.
> (5) A social security number.
> (6) Any other identifier that permits the physical or online contacting of a specific individual.
> (7) Information concerning a user that the Web site or online service collects online from the user and maintains in personally identifiable form in combination with an identifier described in this subdivision.

From California Assembly Bill No. 375 (2017) (specific to broadband privacy):

> Customer personal information means information collected from or about an individual customer or user of the customer's subscription that is made available to the internet service provider by a customer or user of the customer's subscription solely by virtue of the provider-customer relationship, including:
>
> (1) Name and billing information;
> (2) Government-issued identifiers, such as Social Security number;
> (3) Information that would permit the physical or online contacting of an individual, such as physical address, email address, phone number, or IP address;
> (4) Demographic information, such as date of birth, age, gender, race, ethnicity, nationality, religion, or sexual orientation;
> (5) Financial information;
> (6) Health information;
> (7) Information pertaining to minors;
> (8) Geolocation information;
> (9) Information from the use of the service, such as web browsing history, application usage history, content of communications, and origin and destination Internet Protocol (IP) addresses of all traffic;
> (10) Device identifiers, such as media access control (MAC) address and international mobile equipment identity (IMEI); and
> (11) Information concerning a customer or user of the customer's subscription that is collected or made available and is maintained in personally identifiable form

List-type definitions can be combined with linkability standards. A combined definition of PII can be an effective way to protect privacy while providing clearly defined standards for when information can and cannot be used without consent or when its breach triggers notification requirements.

DE-IDENTIFICATION

The boundaries of “personal information” or “personally identifiable information” under various laws continue to provoke vexing policy questions, as noted above. “Anonymized” or “de-identified” information is often exempted from privacy laws, but legislative proposals generally lack any definition of either term.

One approach to defining de-identification is to focus on the outcome, describing the set of characteristics that de-identified information must exhibit before it can be released. Another approach

would be to consider de-identification a process and refer to various statistical methods that may be used to de-identify data to limit identity disclosure.

De-identification as an outcome
From Privacy Technical Assistance Center's *Data De-identification: An Overview of Basic Terms)*:

> De-identification refers to the process of removing or obscuring any personally identifiable information from student records in a way that minimizes the risk of unintended disclosure of the identity of individuals and information about them. Specific steps and methods used to de-identify information (see disclosure limitation method for details) may vary depending on the circumstances, but should be appropriate to protect the confidentiality of the individuals. While it may not be possible to remove the disclosure risk completely, de-identification is considered successful when there is no reasonable basis to believe that the remaining information in the records can be used to identify an individual.
>
> Further, when making a determination as to whether the data have been sufficiently de-identified, it is necessary to take into consideration cumulative re-identification risk from all previous data releases and other reasonably available information, including publicly-available directory information and de-identified data releases from education records as well as other sources. In particular, care should be taken to monitor new releases of de-identified individual-level student data that are released with an attached record code.

De-identification as a process

The process of de-identifying sets of data is often referred to as the process of "disclosure limitation." This term refers to several statistical techniques that minimize the risk of inadvertent and/or unauthorized disclosure of personally identifying information. Considerations of data type and sensitivity can determine what method of disclosure limitation is best suited for a certain data set. Legislators crafting legislation where de-identified data will be released to the public should consider the benefits of various disclosure limitation processes. Once again, the education context proves useful to understanding de-identification as a process.

Again, the Privacy Technical Assistance Center has provided the following disclosure limitation methods:

> **Perturbation** is a disclosure limitation method which involves making small changes to the data to prevent identification of individuals from unique or rare population groups. Data perturbation is a data masking technique in that it is used to "mask" the original values in a data set to avoid disclosure. Examples of this statistical technique include swapping data among individual cells to introduce uncertainty, so that the data user does not know

whether the real data values correspond to certain records, and introducing "noise," or errors (e.g., by randomly misclassifying values of a categorical variable).

Suppression is a disclosure limitation method which involves removing data (e.g., from a cell or a row in a table) to prevent the identification of individuals in small groups or those with unique characteristics. This method may often result in very little data being produced for small populations, and it usually requires additional suppression of non-sensitive data to ensure adequate protection of PII (e.g., complementary suppression of one or more non-sensitive cells in a table so that the values of the suppressed cells may not be calculated by subtracting the reported values from the row and column totals). Correct application of this technique generally ensures low risk of disclosure; however, it can be difficult to perform properly because of the necessary calculations (especially for large multi-dimensional tables). Further, if additional data are available elsewhere (e.g., total student counts are reported), the suppressed data may be re-calculated.

GEOLOCATION

Geolocation overview

Many of the devices and services consumers use collect some type of location information. Location information can be particularly sensitive because it can reveal where people go, facilitating inferences about their lifestyle, habits, health status, and relationships. Geolocation information may be more or less precise depending on the type and amount of information collected. Legislation has been introduced at the federal and state level to offer protections for, and restrict access to, various types of location information. Legislative proposals must be tailored to address who may be using location information and how that information is used.

Precise technical definitions of "geolocation" are rare. Geolocation information may be derived from many different technical sources, making it difficult to define. The World Wide Web Consortium (W3C), for instance, explains in its Geolocation API Specification that common sources for geolocation include the Global Positioning System (GPS) and information inferred from network signals such as IP address, RFID, WiFi and Bluetooth MAC addresses, and GSM/CDMA cell IDs, as well as user input. As a result, legislative efforts tend to address geolocation either generally or as derived from specific sources.

Defining and scoping geolocation information

In the context of mobile devices, broad definitions of "geolocation information" may capture collection and use of information that is necessary for certain services, e.g., phone calls, or information directly provided by users via sign-up. As a result, the scope of geolocation information covered by legislation must be carefully delineated, and exceptions may be required to ensure that communications and other services can function.

For example, one legislative proposal in California broadly defined geolocation information to include any data "that can be used to identify the physical location of an electronic device or its user." The bill would have mandated any use of such data to require affirmative express consent by individuals. For example, a clock app on a phone would need permission to check what timezone it is in.

By contrast, another proposal to amend Section 1798.81.5 of the California Civil Code would have defined geolocation information more narrowly:

> Location data generated by a consumer device capable of connecting to the Internet that directly identifies the precise physical location of the identified individual at particular times and that is compiled and retained. "Geolocation information" does not include the contents of a communication or information used solely for 911 emergency purposes.

Precise geolocation

The sensitivity of geolocation data is determined by its precision and the level of detail it contains. For instance, knowing that an individual is frequenting a specific bar in a specific casino is different from determining an individual is visiting the Las Vegas Strip or is located in the state of Nevada.

Because the sensitivity of geolocation data is highly contextual, understanding of what constitutes "precise" geolocation remains particularly contentious.

One self-regulatory organization, The Network Advertising Initiative, offers guidance for determining when location is imprecise (rather than precise). It suggests analyzing data based upon four factors:

- The area of the identified location (e.g., how many decimal places were used in a lat/long coordinate)
- The population density of the located area
- The accuracy of the location data
- The presence and detail of the location's timestamp

The Children's Online Privacy Protection Act, on the other hand, covers geolocation information only to extent that it is "sufficient to identify street name and name of a city or town."

Definitions for Broadband Privacy Legislation

In 2016, the Federal Communications Commission (FCC) promulgated a set of rules to protect the privacy of internet users' personal information collected by broadband providers. Congress passed a resolution repealing the rules in 2017, prompting state legislators to propose bills enacting similar protections at the state level. The following definitions come from acts of Congress and from FCC rules and represent some of the key terms that broadband privacy legislation may need to include.

Broadband Internet Access Service - BIAS (from 2010 Open Internet Order)

CENTER FOR DEMOCRACY & TECHNOLOGY

> "Services that 'provide the capability to transmit data to and receive data from all or substantially all Internet endpoints" as well as "any service the Commission finds to be providing a functional equivalent of the service described in the previous sentence, or that is used to evade protections set forth in these rules."

Customer (from 2016 Broadband Privacy Rule 81 FR 87274)

> For purposes of the rules we adopt today implementing section 222, we define "customer" as (1) a current or former subscriber to a telecommunications service; or (2) an applicant for a telecommunications service.

Sensitive customer personal information (from 2016 Broadband Privacy Rule 81 FR 87274)

> "...sensitive customer PI includes financial information, health information, Social Security numbers, precise geolocation information, information pertaining to children, content of communications, web browsing history, application usage history, and the functional equivalents of web browsing history or application usage history."

For additional information about the key components of the FCC's repealed broadband privacy rules, see CDT's Broadband Privacy Cheat Sheet.

The Center for Democracy & Technology is a nonprofit technology advocacy organization that works to preserve the user-controlled nature of the internet and champion freedom of expression. We support laws, corporate policies, and technology tools that protect the privacy of Internet users, and advocate for stronger legal controls on government surveillance.

1401 K Street NW, Suite 200 Washington, DC 20005