May 9, 2025

De-identification: New Laws, New Threats, New Approaches

Ann Waldo, JD
Principal
Waldo Law Offices

Brian Rasquinha, PhD
Associate Director
Solution Architecture
Privacy Analytics

Patrick Baier, DPhil
HIPAA Privacy Expert
Datavant

Privacy+
Security
Forum

Speaker

Ann Waldo, JD
Principal
Waldo Law Offices, PLLC

Ann Waldo is the Principal in the boutique law firm of Waldo Law Offices in Washington, DC. She provides legal counsel regarding health data privacy, data strategy, and data transactions, as well as public policy and advocacy regarding data privacy. She has worked as Chief Privacy Officer for Lenovo, Chief Privacy Officer at Hoffmann-La Roche, in Public Policy at GlaxoSmithKline, in-house counsel at IBM, and commercial litigation. Ann has a JD from UNC Law School with high honors. She is licensed to practice law in DC and North Carolina and is a member of the Bar of the U.S. Supreme Court. She is passionate about health data, de-identification, and innovation.

Speaker

Privacy+ Security Forum

Brian Rasquinha, PhD
Associate Director, Solution Architecture
Privacy Analytics

Dr. Rasquinha has been working in data de-identification, anonymization, and privacy program advisory since 2017, with particular focus on scaling data privacy programs and on de-identification of unstructured data like text, images, and audio. His focus is on close collaboration between legal, governance, and business groups to design and implement effective de-identification and anonymization solutions. Brian holds a PhD in Mechanical Engineering from Queen’s University, and has contributed to research in statistical shape models for biokinematics, biomechanics, surgical navigation, and surgical robotics. He has previously worked with ultrasound monitoring of thermal ablation at Sunnybrook Research Institute.

Speaker

Patrick Baier, DPhil
HIPAA Privacy Expert
Datavant

Dr. Baier is a mathematician and cryptographer who has been working in statistical disclosure limitation, records linkage and the development of statistical and cryptographic privacy preserving technologies since 2003. Prior to joining Datavant in 2022, he operated an independent consulting business helping clients in private industry and government with de-identification of personal data and HIPAA compliance.

Patrick holds a DPhil in mathematics (algebraic and differential geometry) from Oxford University, UK, and prior to that has studied mathematics and theoretical physics in Freiburg, Germany, and Cambridge, UK.

DE-IDENTIFICATION AND THE LAW(S)

STANDARDS

- **Play a vital role globally by facilitating communication, innovation, progress**
- Early civilizations developed standardized ways to measure time and space – calendars, clocks, length, weight, etc. Some idiosyncratic (*e*.g., King of England's own arm)
- **Int'l trade and Industrial Revolution made greater standardization essential**

→→Calendars – Roman, Mayan, Egyptian, Islamic, Hebrew, Hindu, Persian… Gregorian calendar finally widely adopted by 19th century, now **the** international calendar standard used WW

→→Distance – Scottish mile longer than English mile – Scottish mile outlawed three times!

- **Strong historical trend toward greater harmonization and standardization**

But de-identification standards? New laws are taking us backward to the realm of inconsistent standards

De-Identification Under State Laws

De-Identification

CA CCPA (Original)

- Original CCPA had a novel definition of “deidentification” that applied to ALL data – and wasn’t at all harmonized with HIPAA standard
- No exception for HIPAA de-ID’d data
- Would have resulted in expensive lawyering, contractual wrangling over risk, delays, costs, litigation risk, etc. and impediments to research

- **CA CCPA (Today)**
- HIPAA De-ID’n definition applies to <u>patient information only</u>
- Patient information = “PHI Plus”
- General CCPA definition applies to all other data

De-Identification

How do other new state privacy laws address de-ID?

20 of the 22 enacted to date have a two-tier structure similar to CA's:

- **HIPAA de-ID'n applies to "PHI Plus" (PHI plus other medical data)**
- **New state-specific de-ID definition applies to all other data**

But 2 of the 22 enacted to date **DO NOT** have a two-tier structure similar to CA's.

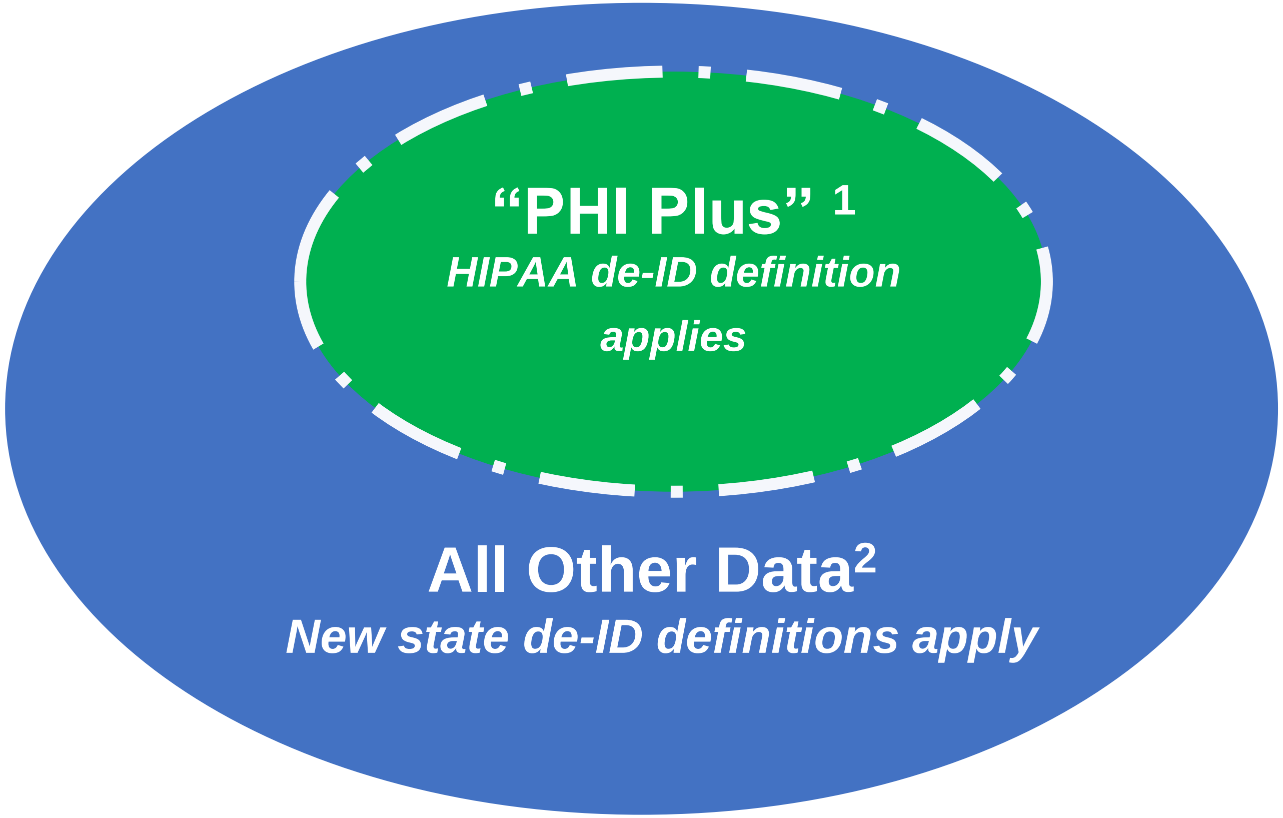
- **In DE and NJ, the new state-specific de-ID definition applies to all data**
- **DE and NJ do not explicitly recognize HIPAA de-ID even for PHI governed by HIPAA**

*Treating WA's and NV's new "consumer health" laws here as general privacy laws due to their breadth of scope

Which state De-ID standard applies to which data? For 20 of the 22 state privacy laws…

[1]**”PHI Plus”** is “patient information” in CA law and has other designations under 13 other state laws. Refers to PHI plus other specified medical data. Examples include PHI, research data subject to Common Rule, Part 2 data, etc. Note – the exact perimeters of what’s included in “PHI Plus” data vary by state.

[2]”**All Other Data**” refers to all data not included in the exemption for “PHI Plus” data. Examples include consumer health data, SDOH, demographic data, etc.

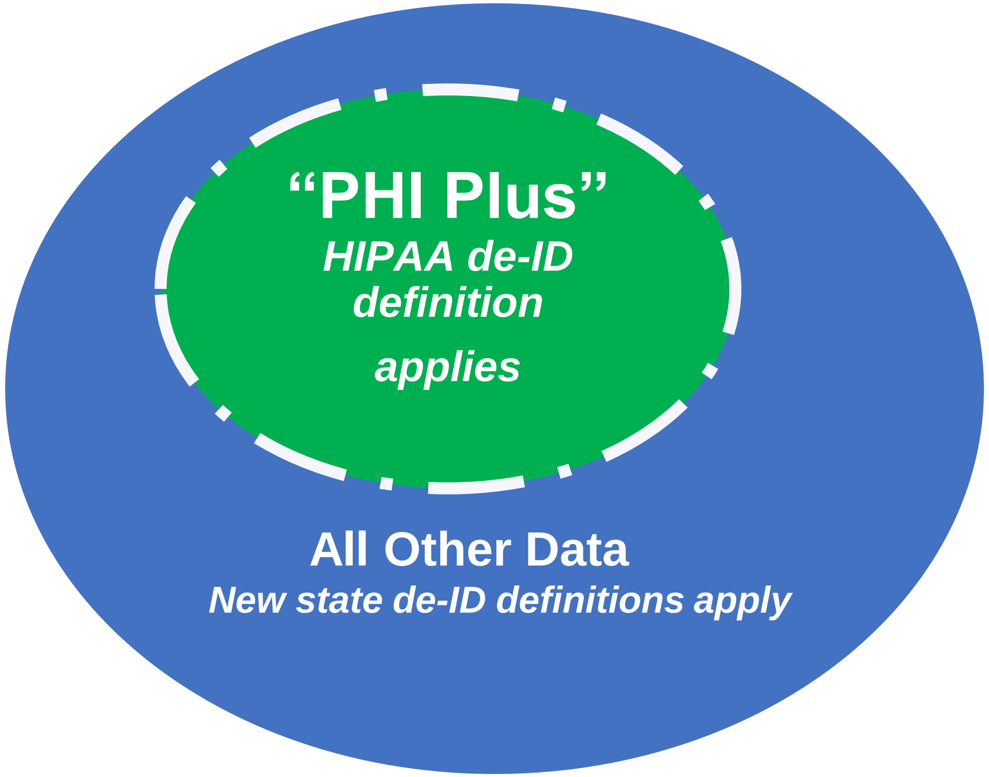


“PHI Plus” [1]
HIPAA de-ID definition applies
All Other Data[2]
New state de-ID definitions apply

More complexities with de-ID’n under the state laws with two-tier de-ID structures

- The perimeter of the inner circle – the “PHI Plus” subject to HIPAA de-ID’n – varies by state
- The de-ID’n language applicable to data in the outer circle varies by state
- Some of the actual definitions include business conduct requirements; some do not

20 of 22 State Privacy Laws



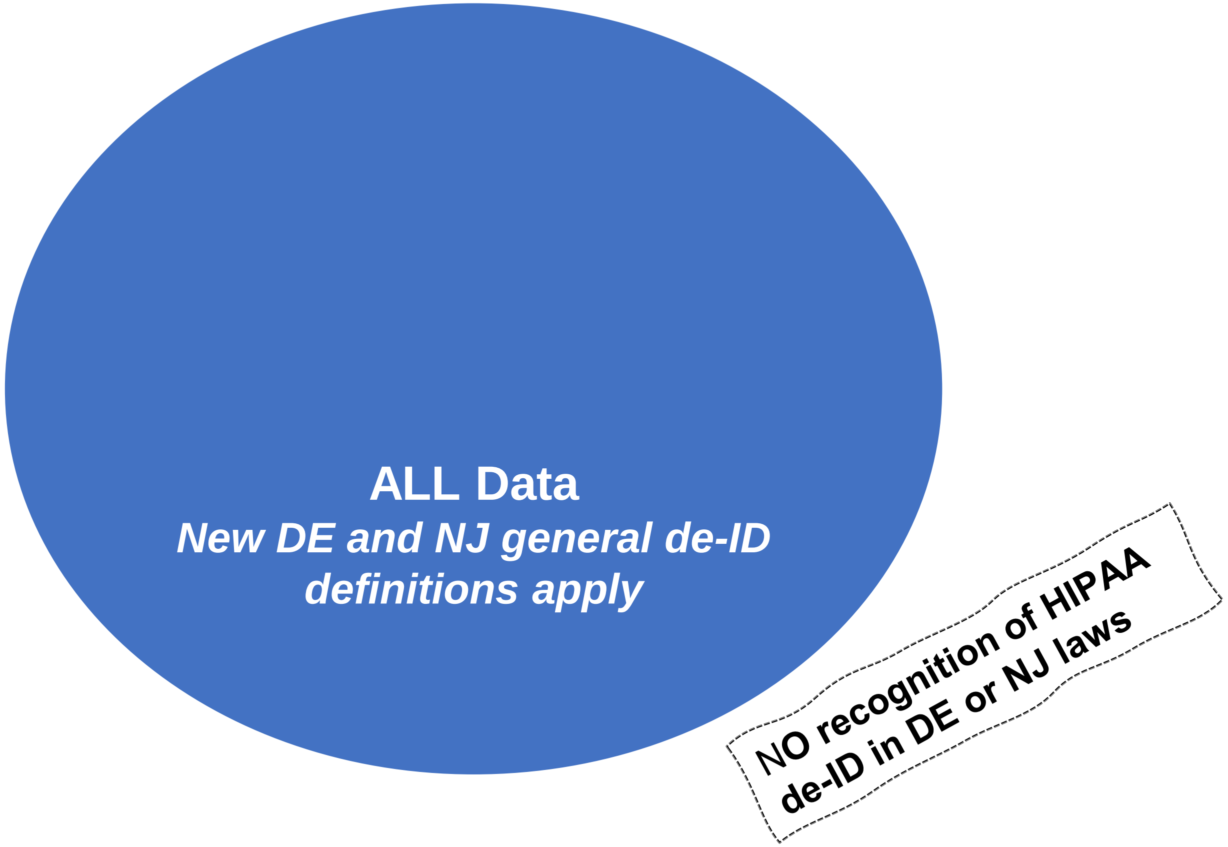
“PHI Plus”

HIPAA de-ID definition applies

All Other Data

New state de-ID definitions apply

DE and NJ Privacy Laws



ALL Data

New DE and NJ general de-ID definitions apply

NO recognition of HIPAA de-ID in DE or NJ laws

Other New State Law Provisions Regarding De-ID’n

1) **CA Ban on re-identification of de-ID’d patient information**

- Cannot re-identify, or attempt to re-identify, de-ID’d patient information (data exempt from CCPA because of newly harmonized de-ID’d definition)
- Exceptions to the ban:
  - TPO under HIPAA (Treatment, Payment, Operations)
  - Public Health under HIPAA
  - Research done in accordance with HIPAA or Common Rule
  - Under a contract to test or validate de-ID’n, provided other uses are banned
  - If required by law

  Note – no other exceptions, including for “white hat” researchers, journalists, etc.

- **Scope - a business or other person ---i.e., broader than the rest of the law’s scope**

De-Identification

Other New State Privacy Provisions re: De-ID

2) Contractual requirements in CA law for sale of HIPAA de-ID’d patient information
3) CA has notice requirements where HIPAA de-ID’d data is sold
4) Multiple states require sellers of de-ID’d data to maintain oversight over customers
5) Pseudonymization makes its first appearance in multiple US laws
6) Data de-ID’n is explicitly a positive factor in DPIAs and risk assessments in some states

De-Identification

Important

- Definitions do not always track well between laws – even within the same state
- If you believe that your data is de-ID’d under a particular state law, such as a data broker law, be sure to check carefully

Potential Consequences

As Divergent Definitions of De-Identification Are Enacted

- FUD – fear, uncertainty, doubt
- Administrative and legal costs
- Delays, friction, contracting obstacles
- Burdens on medical research, medical progress
- Harm to patients and the public

Help educate policymakers about importance of harmonizing de-ID'n

Share best practices re: compliance with de-ID'n standards

Federal Developments Regarding De-Identification

Federal Trade Commission

- Vexed by misrepresentations about data being anonymous when it’s not
- Enforcement actions against companies that claimed data was anonymous when it still could be used by social media or advertisers to re-identify or target individuals
- *Nomi*, *BetterHelp*, *Premom* cases – sharing improperly “anonymized” data despite promises to share only non-identifiable or anonymized data
- 2024 FTC Tech Blog “No, hashing still doesn’t make your data anonymous”

FTC: “Companies often claim and act as if data that lacks clearly identifying information is anonymous, but data is only anonymous when it can never be associated back to a person.

If data can be used to uniquely identify or target a person, it can still cause that person harm.”

FTC: “Claims that data is “anonymous” or “has been anonymized” are often deceptive. Companies may try to placate consumers’ privacy concerns by claiming they anonymize or aggregate data. Firms making claims about anonymization should be on guard that these claims can be a deceptive trade practice and violate the FTC Act when untrue.”

PADFA

Protecting Americans’ Data from Foreign Adversaries Act (2024)

- Unlawful for data brokers to sell, grant access to, etc. **personally identifiable sensitive data** to countries of concern or entities controlled by countries of concern
- Data brokers are entities that, for consideration, make available data about US individuals that the entity did not collect directly from the individuals, subject to a few exceptions
- Sensitive data broadly defined
- **Personally identifiable** data means any sensitive data that identifies or is linked or reasonably linkable, alone or in combination with other data, to an individual or device that identifies or is linked or reasonably linkable to an individual
- Note –This does NOT reference HIPAA de-identified data (whether the data originated from PHI is irrelevant)

DOJ Final Rule - Bulk U.S. Sensitive Personal Data

"Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons"

- **Prohibits**:
  - Certain **transfers***
  - Of ***bulk* US sensitive personal data**
  - And specified **government-related data** (regardless of amount)
  - To **countries of concern**
  - Or **covered persons**
- **Restricts:**
  - **"Restricted transactions"** involving vendor, employment, and investment agreements with countries of concern or covered persons
  - CISA security, recordkeeping, audit, and reporting requirements apply
- **Data brokerage transfers to foreign persons**
  - Onward transfer contractual terms required if transferring to other foreign persons
- **Civil and criminal penalties**

*Many defined terms of art in the Rule – check definitions. Also check exemptions.

DOJ Rule on Bulk Sensitive Data Transfers

Resources

- DOJ Final Rule, Provisions Pertaining to Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons, 28 C.F.R. Part 202, Jan. 8, 2025
- DOJ Fact Sheet, Justice Department Issues Final Rule to Address Urgent National Security Risks Posed by Access to U.S. Sensitive Personal and Government-Related Data from Countries of Concern and Covered Persons, Jan. 8, 2025
- DOJ Press Release, Justice Department Implements Critical National Security Program to Protect Americans' Sensitive Data from Foreign Adversaries, April 11, 2025
- Cobun Zweifel-Keegan,A view from DC: Ready for new US restrictions on nearly all foreign access to personal data?, Jan 3. 2025
- DOJ Releases FAQs and Compliance Guidance for Final Rule Restricting Flow of Bulk Sensitive Personal Data to China and other Countries of Concern, Ropes and Gray, April 14, 2025
- Other presentations at Privacy & Security Academy

Relevance of DOJ Rule to De-Identification

Sensitive personal data explicitly includes anonymized, pseudonymized, de-identified, and encrypted data

Why? National security law, not privacy law

De-identification should always be considered in context

De-Identification –

New Threats,
New Approaches

AI and De-identification – Threats

AI boom fueling changes in how organizations work with data

Advancing methods to attempt re-identification

- More flexible matching
- More inference
- Reduced effort for legacy re-id approaches

Evolving demands for de-identified data

- Increasingly multi-domain
- Increasingly multi-modal (text and more)
- More exploratory datasets

Concerns around generative and agentic models

- Disclosing training data
- Storing/processing query data

Role of governance has even more emphasis!

AI and De-identification – Benefits

AI boom fueling changes in how organizations work with data

Handling identifiers in unstructured data

- Significant detection advancements for text, images, audio, video
- Potential for more realistic surrogates while transforming data

Improved cataloguing tools

- Increased ability to
  - trace data/determine lineage,
  - categorically classify files
  - …and more?

Strong statistical processes can provide validation of AI tools

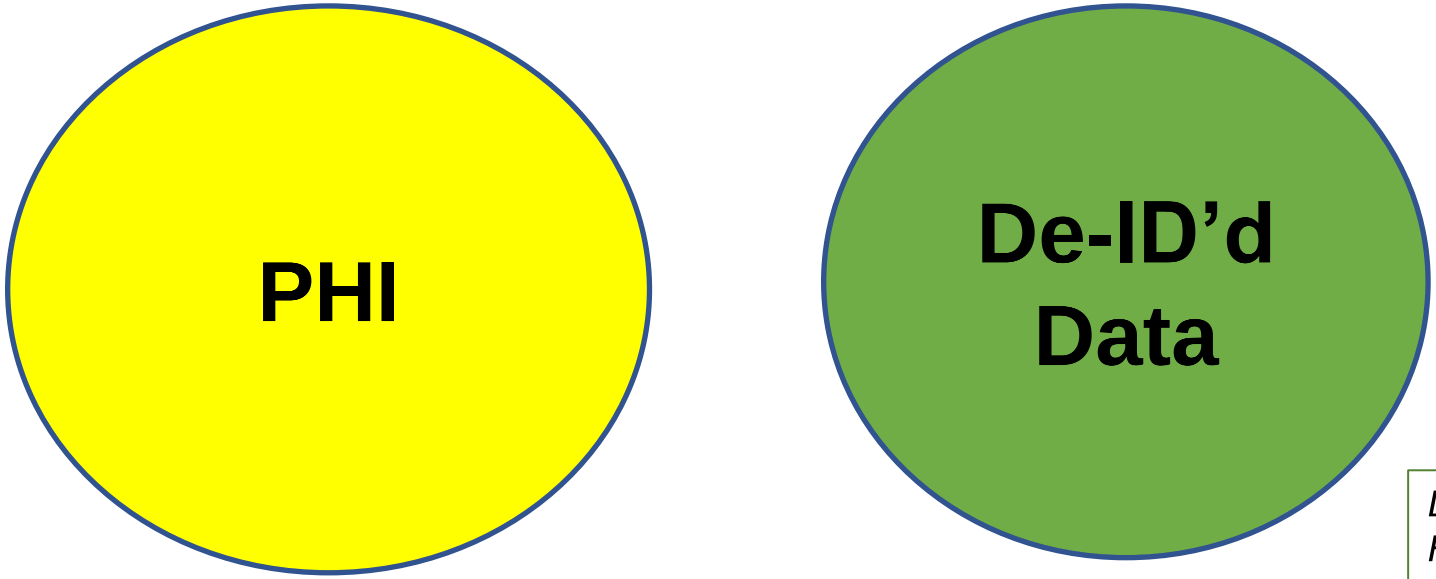
PRIVACY ANALYTICS
an IQVIA company

Questions

Reference Slides

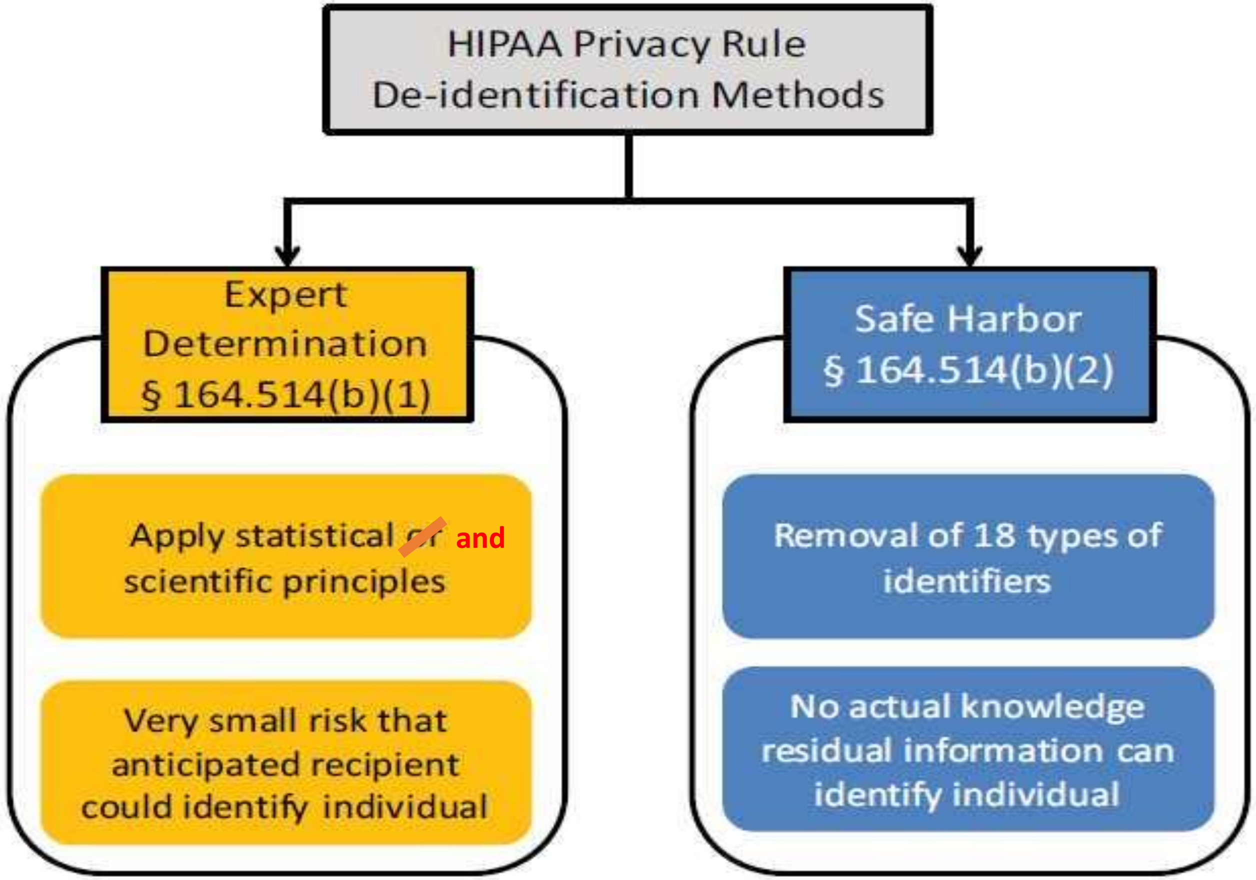
De-identification under HIPAA - Basics

Sharp legal divide in HIPAA between de-identified data and PHI



De-ID’d data is outside HIPAA
HHS has no jurisdiction
Contract restrictions may apply

Two Methods of HIPAA De-identification



HIPAA Privacy Rule
De-identification Methods

Expert Determination
§ 164.514(b)(1)

Apply statistical ~~or~~ and scientific principles

Very small risk that anticipated recipient could identify individual

Safe Harbor
§ 164.514(b)(2)

Removal of 18 types of identifiers

No actual knowledge residual information can identify individual

Source: HHS Office for Civil Rights (OCR) De-Identification Guidance (November 2012)
[Corrected to match wording of §164.514(b)(1)]

HIPAA §164.514(b)(1) “Expert Determination”

Health Information is not individually identifiable if:

A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:

(i) Applying such principles and methods, determines that the *risk is very small* that *the information could be used*, alone or *in combination with other reasonably available information, by an anticipated recipient to identify an individual* who is a subject of the information; and (ii) Documents the methods and results of the analysis that justify such determination;

HIPAA §164.514(b)(2)(i) -18 “Safe Harbor” Exclusions

All of the following must be **removed in order** for the information **to be** considered **de-identified**.

(2)(i) The **following identifiers of the individual or of relatives, employers, or household members** of the individual, are removed:

(A) Names;
(B) All **geographic subdivisions smaller than a State**, including street address, city, county, precinct, zip code, and their equivalent geocodes, **except for the initial three digits of a zip code** if, according to the current publicly available data from the Bureau of the Census: (*1*) The geographic unit formed by combining all zip codes with the same three initial digits contains **more than 20,000 people**; and (*2*) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.
(C) **All elements of dates (except year)** for dates directly related to an individual, including **birth date**, **admission date**, **discharge date, date of death**; and **all ages over 89** and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older;

(D) Telephone numbers;

(E) Fax numbers;

(F) Electronic mail addresses;

(G) Social security numbers;

(H) **Medical record numbers**;

(I) **Health plan beneficiary numbers**;

(J) Account numbers;

(K) Certificate/license numbers;

(L) Vehicle identifiers and serial numbers, including license plate numbers;

(M) **Device identifiers and serial numbers**;

(N) Web Universal Resource Locators (URLs);

(O) Internet Protocol (IP) address numbers;

(P) Biometric identifiers, including finger and voice prints;

(Q) Full face photographic images and any comparable images; and

(R) **Any other unique identifying number, characteristic, or code** except as permitted in §164.514(c)

Example of harmonized de-identification standard (CA)

[Exempt data includes]

(A) Information that meets both of the following conditions:

(i) It is deidentified in accordance with the requirements for deidentification set forth in Section 164.514 of Part 164 of Title 45 of the Code of Federal Regulations.

(ii) It is derived from patient information that was originally collected, created, transmitted, or maintained by an entity regulated by the Health Insurance Portability and Accountability Act, the Confidentiality Of Medical Information Act, or the Federal Policy for the Protection of Human Subjects, also known as the Common Rule.

Example of a new general de-identification definition (CO)

"De-identified data" means data that cannot reasonably be used to infer information about, or otherwise be linked to, an identified or identifiable individual, or a device linked to such an individual, if the controller that possesses the data:

(a) Takes reasonable measures to ensure that the data cannot be associated with an individual;

(b) Publicly commits to maintain and use the data only in a De-identified fashion and not attempt to re-identify the data; and

(c) Contractually obligates any recipients of the information to comply with the requirements of this subsection (11).

Other New State Provisions Regarding De-ID’n

2) CA contractual requirements for sales of de-ID’d patient information

- A contract for the sale or license of de-ID’d patient information must include the following (or substantially similar) terms:
    - Statement about inclusion of de-ID’d patient info
    - Ban on re-ID’n and attempted re-ID’n
    - Downstream contractual terms that are same or stricter
- Scope - one of the parties resides or does business in CA

Other New State Provisions Regarding De-ID’n

3) CA Privacy Notice Requirements

- Scope - a business (per CCPA)
- If a business sells or discloses de-ID’d patient information that’s exempt from CCPA because of the newly harmonized de-ID’d definition for health data, then it must include in its Privacy Policy:

  (a) a statement that it sells or discloses de-ID’d patient information, and

  (b) whether it uses one or more of:
  the HIPAA Safe Harbor method, or
  the expert determination method.

Other New State Provisions Regarding De-ID’n

4) CA - Applicable Law Applies to Re-ID’d Data

- Scope - a business (per CCPA)
- Data that was exempt from CCPA because it qualified for the newly harmonized de-ID’d definition for patient information, *but then became re-identified,* becomes subject to applicable privacy law, including HIPAA, CA CMIA, or CCPA, if applicable

Other New State Provisions re: De-ID

5) Pseudonymization makes its first appearance in US law

- Several states now define pseudonymization *a la* GDPR
- If data is properly pseudonymized, certain state obligations don’t apply.
- And some new requirements apply to pseudonymized data
- *Again – the problem is inconsistency – not all new state laws recognize pseudonymization at all*

Other New State Provisions Regarding De-ID’n

6) Multiple States – New Oversight Duties

- Controller that discloses de-ID’d data must:
  - Exercise reasonable oversight to monitor the data recipients’ compliance with contractual commitments re: the data
  - Take appropriate steps to address any breach of the contractual commitments

- *Some states apply these oversight duties only to de-ID’d data; some to both de-ID’d and pseudonymized data*

Other New State Provisions Regarding De-ID’n

7) Multiple States – Benefits of De-ID’d Data

- Some states allow the use of de-ID’d data to be a factor taken into account in Data Protection Assessments
- Some states have this provision for both de-ID’d and pseudonymized data; some just de-ID’d data