TITAN CAPITAL FORENSICS LLC
Office of Statutory Clearance & Quantitative Forensic Engineering • Albuquerque, New Mexico, USA

STATUTORY METHODOLOGY TECHNICAL NOTE · TITAN-WP-2026-12
Official CERN Zenodo DOI: **10.5281/zenodo.24161859** • Version 1.0.0 • 2026-11-12

Client-Side Zero-Egress PHI De-Identification Verification and Cryptographic Audit-Trail Sealing Under HIPAA 45 CFR § 164.514(b) Safe Harbor

Complete 18-Identifier Transformation Lattice, Census ZCTA3 20,000-Population Suppression, Longitudinal Interval Preservation, and Non-Derived Re-Identification Barriers

Shady Ismail (ORCID: 0009-0005-3920-2333)
Founder & Managing Partner, TITAN Capital Forensics LLC • Albuquerque, New Mexico, USA
Practice Division: Clinical Privacy, HIPAA Safe Harbor & Healthcare Forensics Desk • Correspondence: audit@titancapitalforensics.com

DOI: 10.5281/zenodo.24161859
D-U-N-S®: 33-195-8002 | **UNGM:** #1299169
Reference Implementation: https://titan-scribeguard.com
Citation License: CC BY-NC-ND 4.0 International

ABSTRACT

Under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule at 45 CFR § 164.514(b)(2), clinical records qualify as Safe Harbor de-identified data exempt from 45 CFR Part 164 restrictions only when all 18 enumerated identifier classes of the individual—and of relatives, employers, or household members—are deterministically removed or transformed, paired with absence of actual knowledge of residual re-identification risk (§ 164.514(b)(2)(ii)). Cloud-based clinical redaction APIs introduce two severe legal-technical vulnerabilities: first, transmitting raw Protected Health Information (PHI) to external cloud endpoints constitutes a regulated disclosure requiring an executed Business Associate Agreement (45 CFR §§ 164.308(b)(1), 164.502(e)) before a single token is redacted; second, legacy pipelines routinely violate strict boundary rules on 3-digit ZIP Code Tabulation Areas (suppressing ZCTA3 prefixes with population <= 20,000 to '000' under § 164.514(b)(2)(i)(B)), age > 89 top-coding including birth-year suppression (§ 164.514(b)(2)(i)(C)), and § 164.514(c)(1)'s prohibition against hash-derived re-identification codes over low-entropy 9-digit SSN spaces (~29.9 bits). This technical note formalizes a client-side zero-egress transformation lattice with zero-PHI cryptographic transformation certificates satisfying 45 CFR § 164.530(j).

Keywords: HIPAA Safe Harbor | 45 CFR 164.514(b) | PHI De-Identification | ZCTA3 20000 Population Rule | Age 90 Top-Coding | 45 CFR 164.514(c) Re-Identification | Zero-Egress WebAssembly | Clinical Audit Trails

Classifications: JEL: I18, K32, C88 | ACM CCS: Security and privacy -> Data anonymization and sanitization; Applied computing -> Health informatics

1. Statutory Structure of HIPAA Safe Harbor (`45 CFR § 164.514(a)–(c)`)

Pursuant to `45 CFR § 164.514(a)`, health information is not individually identifiable only if it neither identifies nor provides a reasonable basis to believe it can be used to identify an individual. While `§ 164.514(b)(1)` permits statistical *Expert Determination*, `45 CFR § 164.514(b)(2)` establishes the deterministic **Safe Harbor** standard requiring conjunctive satisfaction of two statutory prongs:

1. **Exhaustive 18-Category Identifier Removal (`45 CFR § 164.514(b)(2)(i)(A)–(R)`):** Removal of names `(A)`; all sub-state geographic subdivisions `(B)` subject to the strict `> 20,000` Census `3-digit` ZIP exception; all sub-year date elements and all ages/years indicative of **age over** `89` `(C)`; telephone `(D)`, fax `(E)`, and email `(F)` addresses; Social Security numbers `(G)`; medical record `(H)`, health plan beneficiary `(I)`, account `(J)`, and certificate/license `(K)` numbers; vehicle `(L)` and device `(M)` serial identifiers; URLs `(N)` and IP addresses `(O)`; biometric identifiers `(P)`; full-face photographs and comparable images `(Q)`; and any other unique identifying number, characteristic, or code `(R)`—except as permitted by `§ 164.514(c)`.
2. **Absence of Actual Knowledge (`45 CFR § 164.514(b)(2)(ii)`):** The covered entity does not have actual knowledge that the remaining information (such as rare occupation titles, genomic sequences, or unredacted free-text clinical narratives mentioning relatives or employers) could be used alone or in combination to identify the subject.
3. **Prohibition on Identifier-Derived Linkage Codes (`45 CFR § 164.514(c)(1)–(2)`):** As explicitly affirmed in `HHS Office for Civil Rights (OCR) De-Identification Guidance`, publishing an unsalted cryptographic hash $\mathrm{SHA256}(\mathrm{SSN})$ or $\mathrm{SHA256}(\mathrm{MRN})$ inside a de-identified dataset **violates** `§ 164.514(c)(1)`, because the code is *derived from* information about the individual and is trivially invertible over finite identifier spaces.

2. Formal Transformation Lattice & Entropy Bounds

2.1 Exact Boundary Operators on `ZCTA3` Population (`(B)`) & Age Top-Coding (`(C)`)

Let $z = d_1d_2d_3d_4d_5$ denote a 5-digit U.S. Postal ZIP code with 3-digit prefix $z_3 = d_1d_2d_3$, and let $P_{\text{Census}}(z_3) \in \mathbb{Z}_{\geq 0}$ denote the current publicly available U.S. Bureau of the Census combined population across all ZIP codes sharing prefix z_3. Notice that `45 CFR § 164.514(b)(2)(i)(B)` specifies a **strict inequality** (*"contains more than 20,000 people"*—so an area of exactly `20,000` inhabitants must still be suppressed!):

$$\mathcal{T}_{\text{ZIP}}(z) = \begin{cases} z_3 & \text{if } P_{\text{Census}}(z_3) > 20{,}000 \\ \texttt{"000"} & \text{if } P_{\text{Census}}(z_3) \leq 20{,}000 \ \vee \ |z| < 3 \end{cases}$$

Similarly, let $t_{\text{dob}} = (y_b, m_b, d_b)$ denote birth date and $t_{\text{enc}} = (y_e, m_e, d_e)$ denote a clinical encounter date, with exact integer age $a(t_{\text{dob}}, t_{\text{enc}}) \in \mathbb{Z}_{\geq 0}$. Under `45 CFR § 164.514(b)(2)(i)(C)`, whenever $a > 89$ (`a >= 90`), retaining y_b alongside y_e would immediately disclose $y_e - y_b \geq 90$ ("elements of dates **(including year)** indicative of such age"); therefore the joint date-and-age morphism $\mathcal{T}_{\text{temporal}}$ must suppress y_b whenever $y_e - y_b > 89$ or $a > 89$:

$$\mathcal{T}_{\text{temporal}}(t_{\text{dob}}, t_{\text{enc}}) = \begin{cases} (\text{age} = a,\ \text{birth_year} = y_b,\ \text{enc_year} = y_e) & \text{if } a \leq 89 \ \wedge \ (y_e - y_b) \leq 89 \\ (\text{age} = \texttt{"90+"},\ \text{birth_year} = \perp,\ \text{enc_year} = y_e) & \text{if } a > 89 \ \vee \ (y_e - y_b) > 89 \end{cases}$$

Crucially, because longitudinal clinical utility (such as hospital length of stay $\Delta t_{\text{LOS}} = t_{\text{discharge}} - t_{\text{admit}}$ in days or readmission intervals) is destroyed if raw day/month fields are stripped *before* interval extraction, computing exact non-identifying relative integer durations $\Delta t_{u,v} = \text{epoch_days}(t_v) - \text{epoch_days}(t_u)$ inside client-side memory prior to stripping (m, d) preserves `100%` of relative clinical kinetics without disclosing calendar dates (provided inter-event intervals do not cross-identify rare external events under `(b)(2)(ii)`).

2.2 Dictionary Inversion Proof Against Unsalted Hashes (`45 CFR § 164.514(c)(1)`)

The entire U.S. 9-digit Social Security Number domain satisfies $|\mathcal{D}_{\text{SSN}}| < 10^9$, yielding Shannon/Hartley upper-bound entropy:

$$H_0(\mathcal{D}_{\text{SSN}}) = \log_2(10^9) \approx 29.90 \text{ bits}$$

Evaluating 10^9 `SHA-256` hashes requires `< 0.15 seconds` on a commodity GPU (`> 10 GH/s`), allowing `100%` deterministic inversion of unsalted SHA256(SSN) pseudonyms. Compliant longitudinal linkage under `§ 164.514(c)(1)-(2)` requires either: (i) a non-derived `128-bit` / `256-bit` CSPRNG random token $u_p \xleftarrow{\text{R}} \{0,1\}^{256}$ stored exclusively inside the Covered Entity's air-gapped cross-walk vault, or (ii) enclave-confined `HMAC-SHA256` $(K_{\text{HSM}}, \text{ID}_p)$ where the `256-bit` hardware-isolated key K_{HSM} never leaves the Covered Entity and is never disclosed (`§ 164.514(c)(2)`).

3. Taxonomy of HIPAA Safe Harbor Pipeline Violations

Failure Class	CFR Anchor	Root Privacy / Legal Defect	Deterministic Client-Side Gate
F-01: Pre-BAA Cloud API PHI Egress	`45 CFR §§ 164.308(b)(1), 164.502(e)`	Streaming unredacted clinical notes or HL7/FHIR JSON to external SaaS/LLM endpoints to perform redaction.	`100%` local browser/workstation `Wasm` execution (`connect-src 'none'` ; `0` raw PHI network packets).
F-02: ZCTA3 `<= 20,000` Boundary Leak	`45 CFR § 164.514(b)(2)(i)(B)`	Truncating all 5-digit ZIP codes to `ZIP[0:3]` without suppressing low-population prefixes (`036` , `059` , `821` , `890` at `<= 20k`).	Embedded Census ZCTA3 lookup enforcing strict `pop > 20000` predicate else `"000"` .
F-03: Nonagenarian Birth-Year Leak	`45 CFR § 164.514(b)(2)(i)(C)`	Masking `"Age: 93"` to `"90+"` while leaving `"DOB: 1933"` or `"93-year-old"` in structured or narrative fields.	Coupled age/birth-year algebraic clamp (`age > 89 => birth_year = null`) + clinical lexical scanner.
F-04: Unsalted SHA-256 Pseudonymization	`45 CFR § 164.514(c)(1)` & `OCR Guidance`	Replacing `SSN` or `MRN` with `SHA256(SSN)` (`29.9-bit` brute-force keyspace, legally prohibited derived code).	CSPRNG non-derived UUIDv4 crosswalk or local HSM-held `256-bit` `HMAC-SHA256` barrier.

4. Zero-PHI Cryptographic Transformation Sealing (`45 CFR § 164.530(j)`) & Verification Summary

A subtle compliance trap arises when logging de-identification proofs for the `6-year` HIPAA documentation retention mandate (`45 CFR § 164.530(j)`): if the audit log records `"Replaced 'John Doe' at offset 42"` , **the audit log itself becomes regulated PHI**!

In the reference verification implementation at `titan-scribeguard.com` , all 18-identifier redaction, ZCTA3 population suppression, and nonagenarian top-coding rules execute locally in client-side WebAssembly (`0` PHI network egress):

1. **In-Memory Clinical Interval Extraction Prior to Date Truncation:** Relative clinical durations (Δt_{LOS} in days, readmission intervals) are computed inside isolated `Wasm` linear memory prior to stripping month/day tokens, preserving longitudinal research utility while outputting strictly year-level (`<= 89`) or top-coded (`"90+"` , `birth_year = null`) Safe Harbor records.
2. **Embedded Census ZCTA3 Strict-Inequality Suppression:** Every 5-digit ZIP code is evaluated locally against embedded Census `ZCTA3` population aggregates under the strict `pop > 20,000` predicate, mapping all `<= 20,000` regions deterministically to `"000"` .
3. **Zero-PHI Cryptographic Audit Certificate:** The output compliance certificate logs strictly **zero-PHI structural counters**—transformed identifier counts across `(A)-(R)` , Census table SHA-256 digest, salted input/output commitments, and `RFC 8785` canonical parameters—sealed locally with `Ed25519` (`RFC 8032`).

Verification Tier	Clinical & Identifier Input Vector	Formal / Empirical Outcome
Tier A: Realistic Encounter	Patient born `1962` , encounter `2026-11-12` , ZIP `87102-1234` (`ZCTA3 871` pop `685,000 > 20,000`), `5-day` inpatient stay.	Outputs `ZCTA3 = "871"` , `age = "64"` , `birth_year = "1962"` , `enc_year = "2026"` , exact `LOS = 5` days; `0` bytes network egress.
Tier B: Exact Statutory Cliffs	`(i)` `ZCTA3` population `20,000` vs. `20,001` ; `(ii)` Birth year `1937` (`age 89`) vs. `1936` (`age 90` in `2026`).	`(i)` `20,000 -> "000"` (`strict > 20,000`), `20,001 -> "890"` ; `(ii)` `1937 -> ("89", "1937")` vs. `1936 -> ("90+", null)` .
Tier C: Adversarial Hash Inversion	Unsalted `SHA256(SSN)` over 9-digit SSN space (`29.90 bits` entropy) vs. enclave-confined `256-bit` `HMAC-SHA256` / CSPRNG UUID.	Unsalted `SHA-256` inverted via dictionary sweep (`violates § 164.514(c)(1)`); non-derived / `256-bit` keyed barrier resists inversion.

5. Primary Statutory & Normative References

1. U.S. Department of Health and Human Services (`HHS`). *Standards for Privacy of Individually Identifiable Health Information (HIPAA Privacy Rule)*, codified at `45 CFR § 164.514(a)-(c)` and `45 CFR § 164.530(j)` .
2. U.S. Department of Health and Human Services, Office for Civil Rights (`OCR`). *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*.
3. U.S. Department of Health and Human Services (`HHS`). *HIPAA Security Standards for the Protection of Electronic Protected Health Information*, `45 CFR §§ 164.308(b)(1)` and `164.312(b)-(c)` .
4. U.S. Bureau of the Census. *ZIP Code Tabulation Areas (`ZCTAs`) Decennial Population Tables for 3-Digit Prefix Aggregation*.
5. Garfinkel, S. L. (2015). *De-Identification of Personal Information*. National Institute of Standards and Technology (`NIST`) `NISTIR 8053` .

REFERENCE VERIFICATION IMPLEMENTATION

The formal mathematical invariants, deterministic state-machine rules, and cryptographic attestation specifications documented in this technical note correspond to the reference verification implementation maintained by **TITAN Capital Forensics LLC** (Albuquerque, New Mexico, USA • ORCID: 0009-0005-3920-2333) at **https://titan-scribeguard.com**. Published strictly as an open-citation scholarly methodology specification under *Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)*. Zero source code or executable software distributed.