WP-2026-004 — XyDromatics De-Identification Engine: Secure Cross-Network Ingest, De-Identified at the Trust Boundary (Whitepaper)

Document number: DOC-2026-442

Whitepaper: WP-2026-004

Version: 5

Category: Marketing Material

Product family: fleet (XyDromatics De-Identification Engine)

Owner: Synthology Healthcare Solutions Group, LLC (SHSG)

Audience: Hospital / IDN CISOs, Security & Privacy Officers, Research-IT leadership, Imaging / PACS directors, research-data governance

Regulatory posture: Non-device software under **FD&C Act §520(o)(1)(D)**. No clinical or diagnostic function; nothing in this whitepaper describes a medical device. De-identified output is **Research Use Only (RUO)**.

Companion: Renders on the marketing site at `/whitepaper/deid-secure-ingest`

Cross-references: RI-2026-353 (secure cross-network ingest); RI-2026-347 (coded-concept cohort builder)

Executive summary

Research and AI programs need imaging and reports moved into a research environment — but that data starts life as PHI, and the network path between the clinical source and the research destination is exactly where exposure risk concentrates. The XyDromatics De-Identification Engine resolves this by treating de-identification as a **trust boundary**, not a downstream cleanup step: data is accepted only over authenticated, encrypted, fail-closed channels, de-identified in place at the point of entry, and forwarded so that **only de-identified output ever leaves the engine.**

- **Three secure ingest channels** — DICOM-TLS C-STORE, MLLP-over-TLS for HL7 v2 reports, and DICOMweb STOW-RS — each with TLS, and SHA-256 client-certificate thumbprint pinning **available for mutual authentication (enforced when enabled).** DICOMweb enforces client-certificate authentication by default; C-STORE and MLLP add mutual-TLS pinning on operator opt-in. **All three channels are fail-closed and off by default.**
- **De-identification at the boundary** — HIPAA Safe Harbor tag transforms (45 CFR §164.514(b)(2)) whose residual re-identification risk is certified under a customer **§164.514(b)(1) Expert Determination**, plus keyed pseudonymization, interval-preserving date shift, and a fail-closed content gate for free text, burned-in pixels, and encapsulated documents.
- **The re-identification key never crosses the wire and never persists in the cloud.** De-identified output carries no key and cannot be re-identified downstream.

- **Two deployment models, one trust boundary** — run the engine in Synthology's cloud or entirely on-premises; the guarantee is identical.

This paper covers the ingest security model, the de-identification assurance level, what does and does not leave the engine, how it compares to the usual alternatives, and the regulatory and privacy posture. It describes *what* the engine guarantees and the *assurance level* — not sensitive internal key mechanics.

---

§1 — The problem

1.1 The research-data movement gap

Every imaging AI pilot, registry, and multi-site study begins the same way: someone has to get studies and their associated reports out of the clinical environment and into a research one. The clinical environment is governed by HIPAA and hospital security policy; the research environment — a cloud tenant, an external reading service, an AI vendor — is not the covered entity's protected network. The data has to cross that gap.

1.2 What happens today

In practice, that crossing is where controls thin out. Studies are exported to a share and copied by hand. Reports are forwarded over plaintext HL7 MLLP inside "the trusted network" that turns out to extend further than assumed. De-identification, when it happens, happens *after* the data has already traversed one or more hops as PHI — or it is applied inconsistently across the imaging channel and the report channel, so a de-identified image and its de-identified report can no longer be lined up without a re-identification table that itself becomes a liability.

1.3 Downstream consequences

The result is a set of coupled failures a CISO recognizes: PHI on the wire in cleartext; a correlation table that must be secured, retained, and eventually destroyed; de-identified datasets that are internally inconsistent between modalities; and no single, auditable point where "PHI in, de-identified out" can be proven.

1.4 Why it stays unfixed

The gap persists because the two hard problems — moving data securely across a network, and de-identifying it reversibly enough to be useful but irreversibly enough to be safe — are usually solved by different tools at different points in the pipeline. The engine collapses them into one boundary.

---

§2 — The solution

2.1 The core idea: de-identify at the trust boundary

The engine is a store-and-forward, data-processing utility. It accepts imaging (DICOM) and reports (HL7 v2 ORU) at its ingress, de-identifies each object *in place at the point of receipt*, runs a content gate that inspects free text, pixels, and embedded documents, and only then forwards the de-identified result to a research

destination. Identified data is de-identified before it is ever forwarded; nothing that could not be fully de-identified is forwarded at all.

2.2 Why the ingress is the right place

Placing the boundary at ingress means there is exactly one line to audit and one line to trust. Upstream of it, the data is PHI on a controlled, encrypted, authenticated channel. Downstream of it, the data is de-identified RUO output. There is no intermediate stage where partially-cleaned PHI sits on a share or crosses another hop.

2.3 Fail-closed by construction

Every place PHI could leak, the engine refuses rather than degrades. A de-identification failure, an object the content gate cannot clean, an unknown gate action, or a storage failure all cause the object to be held or the transfer to fail — the source keeps the object and retries. "Accept and hope" is never a code path.

---

§3 — Secure ingest: three secure, mutually-authenticatable channels

All three channels share the same security disciplines: transport encryption via TLS 1.2/1.3, SHA-256 client-certificate thumbprint pinning available for mutual authentication, and fail-closed behavior on any credential or certificate-load failure. Mutual-TLS pinning is available on all three and **enforced when enabled** — the DICOMweb channel enforces client-certificate authentication by default, while the C-STORE and MLLP channels add mutual-TLS pinning on operator opt-in (TLS alone yields an encrypted channel; turning pinning on makes it mutually authenticated). **All three channels are off by default** — an on-premises deployment inside an already-trusted network can keep plain DIMSE and plain MLLP unchanged, and turns secure transport on only when data must cross a network boundary.

3.1 DICOM-TLS on the C-STORE listener

The DICOM C-STORE listener can be wrapped in TLS built from the engine's server certificate (PFX or PEM). Mutual TLS is an operator opt-in on this channel; when it is enabled, the peer's certificate is accepted only if its SHA-256 thumbprint appears in the operator-configured pin set — an **empty pin set with mutual TLS on rejects every peer.** If the certificate fails to load, the TLS listener does not come up; there is no plaintext fallback. Independently of TLS, peer identity is gated by the DICOM Calling-AE allow-list, which applies whether or not TLS is in use — so with TLS-encryption-only (pinning not yet turned on), the Calling-AE allow-list remains the peer-identity control. This TLS capability is built on Synthology's shared, independently-verified imaging transport stack.

3.2 MLLP-over-TLS for HL7 v2 reports

The HL7 ORU listener can wrap each accepted MLLP connection in a TLS session that is negotiated **before the first MLLP frame is read.** Mutual TLS is an operator opt-in on this channel; with it enabled, the peer certificate is validated against the same SHA-256 thumbprint pin model. A certificate-load failure means the listener does not start — again, no plaintext fallback. When MLLP-over-TLS is enabled, no report byte crosses a plaintext or unverified channel — the TLS handshake completes before the first frame is read. (With the default-off plain-MLLP listener, report bytes cross an unencrypted channel by design, inside a trusted network; enable MLLP-over-TLS whenever reports leave that boundary.)

3.3 DICOMweb STOW-RS

The engine accepts `POST /dicom-web/studies` (multipart/related Part 10). This channel is guarded by a dedicated fail-closed gate that authenticates **before any DICOM object is parsed**: an unauthenticated request receives `401` up front. This channel is fail-closed on an either/or credential model — a request must present **either** a valid sealed bearer ingest token (compared in constant time) **or** a pinned client certificate (from the TLS handshake, or passed through from a fronting proxy such as Cloudflare Access). Mutual TLS is on by default for this channel, so client-certificate authentication is required by default **unless a sealed bearer ingest token is configured and presented** as the alternative. With neither credential configured, every caller is rejected. Authenticated objects flow into the same receive path the C-STORE listener uses, so they inherit the entire de-identification and content-gate pipeline.

3.4 Verified at the wire

Consistent with our wire-roundtrip verification discipline, each channel is covered by an integration smoke that asserts, **on the receiver side**, that a real object crossed intact — and that a rogue peer, a missing certificate, or a wrong bearer token is rejected fail-closed. Secure ingest is proven at the protocol boundary, not merely in isolation.

3.5 Secrets are sealed, never emitted

Certificate *paths* and pinned *thumbprints* are non-secret configuration. Certificate passwords and the DICOMweb bearer token are reversible secrets: they are sealed at rest inside a whole-config encryption envelope and are **never returned by the API** — administrative surfaces expose only boolean "is-set" flags, never the values.

---

§4 — De-identification at the trust boundary

The engine's de-identification is applied to every object at receipt, mutating it in place, one-way. This section describes the *value and assurance level* — not the internal key derivation.

4.1 Method and assurance

The mechanical engine applies HIPAA **Safe Harbor** transforms across the 18 identifier categories of 45 CFR §164.514(b)(2), executed at **every level of DICOM sequence nesting** so nested PHI cannot hide below a top-level pass. Because the engine adds pseudonymization and interval-preserving date handling that go beyond pure Safe Harbor, the governing certification is a **§164.514(b)(1) Expert Determination**: the residual re-identification risk of the transformed output is certified by a qualified expert. That Expert Determination is **the customer's own determination, by a qualified expert of the customer's choosing — it is not a determination rendered by Synthology.**

4.2 Pseudonymization

Patient and study identifiers are replaced with deterministic keyed pseudonyms under a **per-site key that never leaves the engine.** Deterministic means the same input yields the same pseudonym for the life of the key, so a patient's studies still reassemble downstream — with **no retained mapping table.** The transformation is one-way; the pseudonymous output cannot be reversed without the on-engine key.

4.3 Interval-preserving date shift

Dates are shifted by a per-patient whole-day offset, so **inter-exam intervals and chronological order are preserved exactly while absolute dates are obscured.** Time-of-day is retained. Ages over 89 are handled per Safe Harbor (§164.514(b)(2)(iii)). When a date is genuinely shifted, the object is stamped with the DICOM Attribute Modification method code (DCM 113107, "Retain Longitudinal Temporal Information Modified Dates") and the standard PS3.15 de-identification method attributes — an honest, machine-readable record of what was done.

4.4 VR-aware substitution and private-tag removal

Replacements are always shaped to fit the target DICOM value representation, so a structural field never breaks the run. Private and proprietary (odd-group) tags are removed by default at every level — a belt-and-braces catch for vendor PHI that a named-tag policy alone would miss. Study and instance UIDs are deterministically remapped so studies still reassemble without a retained map; structural UIDs (SOP Class, Transfer Syntax) are deliberately left intact.

4.5 The content gate

Structured tags are not the only place PHI hides. The fail-closed content gate classifies each object and cleans three residual risks, with a per-lane operator policy of **clean / quarantine / reject** — and any unknown or mis-typed action fails **safe to quarantine**:

- **Free-text SR narratives** — scrubbed with a two-pass PHI scrubber (pattern shapes, then known identifiers). Because free-text scrubbing is best-effort, quarantine remains the conservative default.
- **Burned-in pixel text** — redacted by irreversible pixel replacement using a **local, bundled OCR engine running as a subprocess — never cloud OCR.** An object where text cannot be confirmed clean is quarantined, not forwarded.
- **Encapsulated documents** — PDFs are rasterized, OCR-redacted, and rebuilt as image-only documents that **drop the selectable text layer and all document metadata;** unsupported document types fail closed to quarantine.

Quarantined objects are sealed on the engine for operator review and are **never silently forwarded.**

4.6 Cross-protocol alignment — no correlation table

Reports arriving over HL7 and images arriving over DICOM are de-identified so that the **same medical-record number produces the same pseudonym and the same date-shift offset on both paths.** The consequence is decisive: a de-identified report lands on its de-identified study's timeline **with no correlation table and in any arrival order.** There is no lookup table to secure, leak, or destroy — the alignment is a property of the computation, not of a stored map. For HL7, the engine is strictly store-and-forward and returns a positive acknowledgment only after the downstream research sink accepts; any message it could not fully de-identify and forward is NACK'd so the **sender holds and retries.**

---

§5 — What leaves the engine

The boundary guarantee is precise about the **reversibility boundary**: de-identification is non-reversible *downstream*. The engine internally holds a §164.514(c) re-identification capability — the per-site key, the per-patient offset, and optional sequential labels — but that capability lives **only on the engine**, sealed at rest.

- **What leaves:** de-identified DICOM (pseudonyms, shifted dates, remapped UIDs) and de-identified HL7 reports, forwarded to the research destination (an RUO de-identified DICOM archive such as XyDromatics Research, a downstream HL7 research sink, and/or external consumers such as AI vendors or outside reading services, which receive pseudonymized studies only).
- **What never leaves:** the re-identification key. It does not travel with the data, is not embedded in the output, and **never crosses the wire.** Because the output carries no key, downstream consumers cannot re-identify it.
- **In the Synthology-hosted deployment:** no PHI and no re-identification key persist in the cloud — **only de-identified output.**

An optional coded-concept cohort capability can additionally populate a PHI-free, pseudonym-keyed store of standards-**coded** clinical concepts (never free-text PHI blobs) for cohort discovery. It is opt-in, is triple-gated (operator enablement, a licensed add-on, and a current §164.514(b)(1) attestation), and any **cohort query resolved through the coded-concept store enforces a k-anonymity floor of k ≥ 11** — a below-k coded-concept cohort is suppressed and cannot be materialized, snapshotted, or exported. (This floor governs the coded-concept query path; it is not a claim about every possible cohort dimension.) The cohort lane is deliberately additive — a harvest error never affects the de-identified image forward.

---

§6 — Two deployment models, one trust boundary

The engine ships in two topologies, and the security guarantee is identical in both.

	Synthology-hosted (cloud)	**Customer-operated (on-premises)**
Where the engine runs	Synthology's cloud, alongside the research archive	Entirely inside the customer's network
How data arrives	Customer pushes over a secure channel (DICOM-TLS / MLLP-TLS / STOW-RS)	Local push; secure channels optional inside the trusted network
Trust boundary	At the engine ingress	At the engine ingress
Re-id key custody	On the engine only; never persists in the cloud	On the engine only, inside customer infrastructure
What persists at the destination	De-identified output only	De-identified output only
HIPAA relationship	Synthology as **Business Associate under a BAA**	Customer-operated; no BAA needed for hosting

Either way, PHI is encrypted and authenticated on arrival, de-identified at the boundary, and only de-identified output moves onward.

---

§7 — Comparison

The engine is best understood against the two ways research data usually crosses the clinical/research gap today.

| | **De-identify at the boundary** (this engine) | **Export-then-clean** (hand export + downstream de-id) | **Plaintext MLLP/DIMSE inside "the trusted network"** |

|---|---|---|

| PHI encrypted in transit to research | ✓ (when secure channels enabled) | ✗ (typically copied/moved as PHI first) | ✗ |

| De-identified before any onward hop | ✓ (at ingress) | ✗ (cleaned after ≥1 PHI hop) | ✗ |

| Single auditable "PHI in, de-identified out" line | ✓ | ✗ (multiple hops, multiple tools) | ✗ |

| Cross-modality alignment without a correlation table | ✓ (computed, not stored) | ✗ (usually needs a re-id table) | ✗ |

| Re-identification key stays on the boundary | ✓ (never crosses the wire) | Varies (often a retained map) | N/A |

| Fail-closed on anything it cannot clean | ✓ | Varies | ✗ |

Where we fit. If your research data never leaves a genuinely trusted network, plaintext DIMSE/MLLP may be acceptable — and the engine supports that (secure channels are off by default). The engine earns its place the moment data must cross a network boundary, an external vendor, or a cloud tenant: it makes ingress the one place PHI is encrypted, authenticated, and de-identified, instead of spreading that responsibility across hand-export steps and a downstream cleanup tool.

7.1 What the engine won't do

- **No clinical read.** It has no diagnostic or interpretive function; it is a data-processing utility, not a viewer or a decision aid.
- **No downstream re-identification.** The re-identification key never leaves the engine, so de-identified output cannot be reversed by any downstream consumer.
- **No guaranteed cohort completeness.** The optional coded-concept harvest lane is intentionally fail-open so a harvest error never blocks the de-identified image forward — which means the coded-concept store is best-effort and can under-populate relative to the archive. It is an additive discovery aid, not a system of record.
- **It does not render the Expert Determination.** The §164.514(b)(1) determination is the customer's own, by a qualified expert of the customer's choosing.

---

§8 — Regulatory & privacy posture

8.1 Classification — §520(o)(1)(D) non-device

The XyDromatics De-Identification Engine is **non-device software under FD&C Act §520(o)(1)(D)**. It transfers and converts medical-device data (it de-identifies DICOM); it has **no clinical or diagnostic function.**

8.2 De-identification method and RUO output

The de-identification method is **HIPAA Safe Harbor tag transforms (45 CFR §164.514(b)(2)) whose residual re-identification risk is certified under a §164.514(b)(1) Expert Determination** — the customer's own determination by a qualified expert, not a determination by Synthology. **De-identified output is Research Use Only.**

8.3 HIPAA / BAA posture

HIPAA obligations are **orthogonal** to FDA-device status. When Synthology hosts the engine, Synthology can act as a **HIPAA Business Associate under a BAA** — a data-protection relationship, entirely independent of the software's non-device classification. An append-only, PHI-free audit trail is the evidence record; every attestation capture and revocation is logged.

8.4 Cybersecurity posture

The controls in §3–§4 are the security posture in practice: encrypted ingest with mutual-TLS pinning available on every channel (enforced when enabled); fail-closed rejection with no plaintext fallback; certificate-thumbprint pinning; reversible secrets sealed at rest and never emitted by the API; local-only OCR (no data sent to any cloud text service); and PHI-free logging throughout (failures record an exception type and de-identified identifiers only).

8.5 A note on the research destination

Where studies land in XyDromatics Research, that archive carries its **own distinct** posture: Research Use Only under 21 CFR §812.2(c)(3), non-device under §520(o)(1)(D), with DICOM PS3.15 Basic Profile anonymization enforced at ingest and no clinical read. Its viewer panes are labeled *NOT FOR CLINICAL DIAGNOSIS OR DECISION MAKING*. The De-Identification Engine and XyDromatics Research are complementary but separately classified; neither method should be attributed to the other.

---

§9 — Summary & how to engage

Moving imaging and reports into a research environment does not have to mean PHI on the wire, a fragile correlation table, or de-identification bolted on after the fact. The XyDromatics De-Identification Engine makes ingest **the** trust boundary: mutually-authenticatable, encrypted, fail-closed channels in; de-identified, RUO output out; the re-identification key held only on the engine and never crossing the wire — hosted by Synthology under a BAA, or run entirely on your own infrastructure, with the same guarantee.

Shipping status (truthful-claims note): DICOMweb STOW-RS ingest, DICOM-TLS on the C-STORE listener, and MLLP-over-TLS on the HL7 ORU listener are **shipped and wire-roundtrip verified.** [Yes] **All three secure channels are off by default;** an on-premises deployment inside a trusted network keeps plain DIMSE/MLLP until secure transport is turned on. Mutual-TLS client-certificate pinning is available on all three and enforced when enabled (DICOMweb by default; C-STORE and MLLP on opt-in). Burned-in pixel and encapsulated-document redaction use a **local** OCR engine, never cloud OCR. Every claim in this whitepaper traces to shipped behavior.

To engage: contact Synthology Healthcare Solutions Group to scope a deployment model (hosted vs. on-premises), review the channel and certificate configuration for your network, and — where hosting applies —

put a BAA in place. Your privacy team retains and executes the §164.514(b)(1) Expert Determination.

Glossary

- **§520(o)(1)(D)** — The FD&C Act software carve-out (21st Century Cures Act §3060) under which software that transfers, stores, converts, or displays medical-device data is **not a medical device.**
- **Expert Determination (45 CFR §164.514(b)(1))** — A HIPAA de-identification method in which a qualified expert certifies that the residual risk of re-identification is very small. Here it is the **customer's** determination.
- **Safe Harbor (45 CFR §164.514(b)(2))** — The HIPAA de-identification method that removes/generalizes 18 categories of identifiers.
- **Pseudonymization** — Replacing an identifier with a deterministic, keyed, one-way surrogate; reversible only with a key held on the engine.
- **Interval-preserving date shift** — Shifting all of a patient's dates by one whole-day offset so intervals and order are preserved while absolute dates are obscured (DICOM method code 113107).
- **k-anonymity (k ≥ 11)** — A floor applied to cohort queries resolved through the coded-concept store: a coded-concept cohort must contain at least 11 distinct subjects to be previewed, snapshotted, or exported; below-k coded-concept cohorts are suppressed. (This floor governs the coded-concept query path, not every possible cohort dimension.)
- **RUO** — Research Use Only; de-identified output is not for clinical use.
- **DICOM-TLS / MLLP-over-TLS / STOW-RS** — TLS-secured DICOM C-STORE, TLS-secured HL7 v2 MLLP, and the DICOMweb HTTP store service, respectively.

Non-device software under FD&C Act §520(o)(1)(D). Not a medical device; no clinical or diagnostic function. De-identified output is Research Use Only. *© Synthology Healthcare Solutions Group, LLC.*

> **Regulatory classification.** All Synthology and XyDromatics software described here is non-device software under the FD&C Act §520(o)(1)(D) software exclusion — not a medical device, and no clinical diagnostic function. Published by Synthology Healthcare Solutions Group, LLC as a controlled document; the authoritative version is the approved record in the Synthology quality-management system.