AI-Restricted Open Standard (AIROS)

Public Framework Overview v1.2

Maintained by VisionZeroAI (Pty) Ltd

Date: November 2025

Contents

1. Executive Summary

AIROS (AI-Restricted Open Standard) defines a global framework for marking, verifying, and enforcing data boundaries between digital assets and artificial intelligence systems. The purpose of AIROS is to ensure that AI models, applications, and data pipelines respect ownership, consent, and governance controls in a verifiable and interoperable way.

This public overview provides a conceptual understanding of AIROS, its layered architecture, and its alignment with emerging regulatory and security frameworks — without exposing implementation details.

2. Context and Rationale

Artificial intelligence systems increasingly ingest vast quantities of digital data without a universal mechanism to determine whether the content is authorized for AI use. This lack of a standardized 'do-not-ingest' signal has led to privacy violations, intellectual property disputes, and regulatory uncertainty.

AIROS addresses this gap by establishing a cryptographically verifiable provenance framework — enabling responsible AI ingestion and protecting human and organizational data boundaries.

3. Vision and Design Principles

AIROS is designed around the following core principles:

• Interoperability First — compatible across vendors and ecosystems.

• Privacy by Design — no need to expose data content for verification.

• Verifiability — cryptographic assurance of authenticity and intent.

• Governance Alignment — harmonized with global AI accountability standards.

4. AIROS Conceptual Architecture

The AIROS architecture is composed of four primary conceptual layers, representing the journey from content creation to AI policy enforcement:

1. Tagging Layer — embeds provenance metadata within or alongside digital assets.

2. Verification Layer — validates authenticity and policy status of tagged assets.

3. Policy Enforcement Layer — ensures AI systems respect declared restrictions.

4. Trust Registry Layer — maintains cryptographic trust anchors and governance data.

5. Provenance and Policy Framework

AIROS provides a universal method to assert and verify the provenance of digital assets. By attaching a verifiable AIROS tag to a file, organizations can signal whether a file is 'AI-Restricted' or governed under specific consent or licensing policies. The verification process ensures the tag's origin and integrity without exposing the content itself.

6. Security and Trust Model

AIROS employs a layered trust model. Each organization operates under a unique cryptographic identity registered within a verifiable trust registry. Verification is performed locally or federated through trust anchors, ensuring authenticity and non-repudiation. The technical methods are deliberately abstracted in this public edition to preserve implementation integrity.

7. Compliance and Governance Alignment

AIROS supports and complements existing AI governance and regulatory frameworks by providing a tangible technical signal for consent, provenance, and restriction enforcement. Key alignments include:

- EU AI Act — reinforces AI risk classification and lawful data sourcing.
- GDPR — enables technical proof of consent boundaries and data subject control.
- NIST AI Risk Management Framework — implements measurable governance controls.
- ISO/IEC 42001 — aligns with AI management system governance principles.

8. Adoption Scenarios

AIROS can be integrated across multiple domains to enhance transparency and trust:

- Enterprises — embed AIROS tagging in document management and cloud workflows.
- AI Vendors — implement ingestion pre-checks for AI-Restricted manifests.
- Regulators — adopt AIROS as a technical enforcement layer for compliance auditing.
- Developers — integrate AIROS concepts through SDKs and metadata frameworks.

9. Governance and Participation

AIROS is initially stewarded by VisionZeroAI (Pty) Ltd, with a roadmap to transition to a multi-stakeholder governance model. Stakeholders are encouraged to participate in framework evolution, share insights, and contribute through open collaboration channels.

10. Call to Collaboration

VisionZeroAI invites AI developers, policymakers, and organizations to collaborate on the AIROS initiative. The objective is to establish AIROS as a trusted open framework that ensures AI systems respect the boundaries of consent, ownership, and policy.

Interested contributors may contact info@visionzero.co.za or engage through the AIROS-Public GitHub repository (https://github.com/VisionAIROS/AIROS-Public).

Appendix A — Conceptual System Flow

The AIROS System Flow is structured around four sequential processes: Tag → Verify → Enforce → Register. Each stage contributes to the overall provenance and compliance assurance lifecycle.

Appendix B — Terminology

• AI-Restricted — A designation indicating that the content must not be ingested or processed by AI systems.

• AIROS Tag — Structured metadata that communicates ownership, consent, and restriction policy.

• Trust Registry — Governance layer managing organizational identities and trust anchors.

• Policy Enforcement — Mechanism ensuring AI behavior aligns with declared content restrictions.

11. Integration Model and Implementation Pathways

AIROS is designed for seamless integration into existing digital and AI ecosystems. Its conceptual model can be embedded across diverse systems without dependency on specific platforms or technologies.

Key integration pathways include:

• AI Systems — perform a provenance check using AIROS metadata prior to data ingestion.

• Enterprise Data Pipelines — embed verification steps before transferring or indexing digital assets.

• Compliance Systems — automate consent and restriction validation for audit trails.

• Security and Privacy Platforms — align AIROS tagging with DLP, CASB, and trust boundary controls.

This flexible design ensures that AIROS complements, rather than replaces, existing security and governance architectures.

12. Policy Framework Examples

AIROS supports a range of policy signals that define how digital assets may be processed by AI systems. These examples are conceptual and represent common governance use cases.

Policy	Description
AI-Restricted	Do not ingest or process this content in AI models or pipelines.
AI-Allowed	Content is authorized for AI use under declared consent or license.
Quarantined	Flagged content requiring human or compliance review before AI processing.

13. Interoperability and Standards Alignment

AIROS complements and extends existing data provenance and content authenticity standards. It is designed to interoperate with frameworks such as the Coalition for Content Provenance and Authenticity (C2PA), W3C verifiable credentials, and ISO/IEC 27001 trust frameworks. By focusing specifically on AI ingestion and governance, AIROS introduces a missing layer of accountability to ensure machine-readable consent boundaries are enforceable across systems.

14. AIROS Governance Roadmap

AIROS is currently maintained by VisionZeroAI (Pty) Ltd as the founding steward. The long-term governance strategy aims to establish AIROS as an open, multi-stakeholder standard involving representation from AI developers, cybersecurity providers, regulators, and academia.

Governance milestones include the formation of an independent standards board, transparent decision processes, and alignment with international data protection and AI ethics frameworks.

15. AIROS Evolution Timeline

The AIROS framework is continuously evolving toward broader adoption and maturity. The following milestones outline the conceptual roadmap for public awareness and industry collaboration.

Version	Milestone
v1.2	Conceptual framework established and published (Public Edition).
v1.5	Pilot integration testing and industry validation under NDA partnerships.
v2.0	Open standard proposal for cross-industry certification and interoperability alignment.

16. Ethical Impact and Societal Benefit

AIROS reinforces ethical AI development by embedding respect for ownership, consent, and governance directly into the data ecosystem. This ensures that the benefits of artificial intelligence are realized without undermining trust or infringing on digital rights.

By promoting transparency, accountability, and provenance verification, AIROS contributes to a safer, more equitable AI landscape where human and organizational boundaries are respected at every stage of automation.

17. Conclusion

AIROS represents a critical step toward trustworthy, verifiable AI data governance. Its public framework invites collaboration across the global AI ecosystem, enabling a future where every piece of digital content carries a verifiable signal of intent and consent.

VisionZeroAI remains committed to developing AIROS as an open, resilient, and responsible standard that unites technology innovation with ethical integrity.