

Zero Trust Data Pillar Implementation in DoW Environments: Collibra + Data X-Ray

Solution Comparison Brief

February 2026 | Prepared by Ohalo | Confidential

Bottom Line

DoW's Zero Trust Strategy requires data governance that operates **inside the enclave, not at the boundary**. The data pillar necessarily cuts across several technology sets in order to comply with DoW's Zero Trust Strategy. As part of the Zero Trust Strategy, RBAC based perimeter security is gone, and ABAC is the standard. The system must **classify and enforce policy at the file level**, inside the enclave, at mission scale.

The Collibra + Data X-Ray stack is purpose-built for exactly this Strategy.

The question for this evaluation is not whether this stack can work in a DoW environment. It is whether DoW moves now or delays while alternatives close the gaps this environment requires.

The Operational Requirement

The DoD Zero Trust Data Pillar is not a single capability. It is thirty-one interlocking activities across seven capability groups, each building on the last, each requiring the one before it to work. Governing sensitive mission data in a dispersed DoW environment means satisfying all of them — not picking three and hoping the auditor doesn't read the rest of the matrix.

Here is what the strategy actually demands:

1. Data Discovery & Classification at Enterprise Scale (Activities 4.1.1, 4.3.1–4.3.5)

Everything starts here. Components must deploy algorithms that scan, analyze, and classify the content inside every file — not metadata sampling, not perimeter inspection — with full semantic awareness using AI and ML. Those algorithms must be registered, auditable, and produce machine-readable outputs that feed the enterprise data catalog. The strategy then requires a phased progression: implement classification tooling with a full rules engine (create, modify, delete, test for collision and compound inconsistency), move from manual tagging to basic automation, and ultimately achieve full AI-driven tagging with zero human intervention across all supported repositories. Unsupported repositories must be identified, inventoried, and evaluated for decommissioning. This is not optional. It is the foundation every other activity depends on.

2. Enterprise Tagging Standards & Governance (Activities 4.2.1–4.2.2)

Classification means nothing without a controlled vocabulary. The Enterprise defines the data dictionary, the ZT-specific label schema, and the interoperability standards — including mandatory DRM overlays — that every Component must implement. Tags must be standardized, machine-readable, and produce equal computation outcomes for any rule across the environment. An enforcement agent must execute actions based on those computational results. This is the common language requirement: without it, every downstream enforcement decision is unreliable.

3. Data Rights Management & Protection (Activities 4.5.1–4.5.5)

DRM and protection tools must be deployed first to high-risk data objects, then expanded to cover all required repositories with no exceptions. Protection mechanisms must be automatically managed to meet FIPS and environment-specific requirements. DRM enforcement must be driven by data tags using Attribute-Based Access Control (ABAC) standards set by the Enterprise. Data at rest must be encrypted in all repositories. At maturity, DRM integrates with AI/ML tooling for automated encryption, rights management, and protection functions. The end state is unambiguous: no data object bypasses the compliance requirement.

4. Data Loss Prevention Enforcement (Activities 4.4.1–4.4.2, 4.6.1–4.6.4)

DLP enforcement points must be identified, deployed, and evolved through three phases: monitor-only mode with standardized logging, transition to active prevention with ZT indicators that facilitate cooperative cyber enforcement, and finally full integration with automated tagging and AI/ML-driven analytics. Business rules for DLP and DRM must be established and coordinated with Cyber Operations. Standardized logging schemas must be enforced at the Component level and fed into the SIEM. The strategy is explicit: the right people access the right data in the right place at the right time, and everything else is blocked.

5. Continuous File & Data Activity Monitoring (Activities 4.4.3–4.4.6)

Components must monitor the most critical classification levels first, then expand to all regulatory-protected data (CUI, PII, PHI), then integrate database monitoring — all feeding analytics into SIEM, XDR, UEBA, and cross-pillar solutions. File integrity monitoring must occur at the data asset and object level. Extended integration sends data-centric intelligence to DLP, DRM, and behavioral analytics platforms. This is not periodic scanning. This is continuous, active monitoring with real-time feeds into the security stack.

6. Software-Defined Storage & DAAS Integration (Activities 4.2.3, 4.7.1–4.7.3, 4.7.6–4.7.7)

SDS policy must be developed and evaluated against existing storage strategies. DAAS access policy must be integrated with SDS using attribute-based fine-grained controls, implemented in an automated fashion, and expanded to protect all data in all applications. SDS tooling must integrate with DLP, DRM, and ML solutions. The end state is a centralized DAAS security approach across the Enterprise that reduces risk and attack surface area.

7. Identity Provider Integration & Interoperability (Activities 4.7.4–4.7.5)

All data security solutions — DLP, DRM, and storage infrastructure — must integrate with the Enterprise Identity Provider via API, LDAP, or SAML. Access attributes and data location attributes must be interoperable across the stack. The end state: data protection and access granted only to authenticated and authorized users, enforced through the identity layer.

Why This Points to Data X-Ray

Read the section above again. Every one of these seven capability groups requires the same foundational operation: knowing what is inside the data, at the file level, at enterprise scale, with full content awareness.

Data X-Ray is purpose-built for exactly this. Our platform performs deep, AI- and LLM-powered content analysis across all unstructured repositories — not metadata sampling, not keyword matching, but full semantic classification that understands context, nuance, and regulatory categories (CUI, PII, PHI, and mission-specific designations). We produce machine-readable, standardized tags that feed directly into the enforcement chain the DoD requires.

Specifically:

- **Discovery & Classification:** Data X-Ray is the registered algorithm. Our engine scans every file, applies enterprise-standard tags, and populates the data catalog with ground-truth classification — satisfying the full automation progression from 4.1.1 through 4.3.5.

- **Tagging Standards:** We operationalize the Enterprise's controlled vocabulary. Configure the standard once; Data X-Ray applies it consistently, automatically, with collision-aware rule management and full audit trails.
- **DRM & DLP Enablement:** Our classification tags are the triggers. Through native integrations with Microsoft Information Protection, encryption tools like Virtru, and direct feeds to DLP/DRM/CASB tooling like Trellex, Data X-Ray provides the data-centric intelligence that enforcement points consume. We identify what needs protecting; the enforcement stack acts on our output.
- **Monitoring & Analytics:** Every scan produces structured, SIEM-ingestible audit logs — what was found, where, when, and what classification was applied. This is the continuous monitoring feed the strategy demands.
- **ABAC & Identity Integration:** Data X-Ray publishes file-level resource attributes that drive attribute-based access control decisions, integrating with the Enterprise IdP and policy infrastructure.
- **Air-Gapped Deployment:** The platform deploys inside the enclave with zero external connectivity, zero cloud dependency, and zero attack surface introduced by the deployment itself.

The DoD Zero Trust Data Pillar is thirty-one activities that all start with the same question: *what is in the data?* Data X-Ray answers it — at scale, inside the enclave, with the accuracy and automation the strategy requires to move from Target Level to Advanced Level ZT.

Every platform in this evaluation should be measured against all thirty-one. Not four.

Requirements Mapping

ZTA Requirement	Collibra + Data X-Ray
Air-gapped / on-prem deployment	Containerized RHEL/Podman. Zero external connectivity. BYO-LLM for AI inference. Operational in hours. No licensing callbacks. No cloud dependency at any stage. In a tactical environment, deployment latency is a mission liability. This stack eliminates it.
Full-content classification of unstructured files	GenAI-powered classification at 100,000+ words per second. Content and context, not keywords or regex. CUI, PII, ITAR, and mission-specific attributes identified at the file level. Designed for 150M+ file environments.
ICAM / ABAC integration	File-level attributes published directly to Policy Enforcement Points via REST and SCIM. Native integrations to Ping Identity, Virtru, Splunk, and Trellex. Every file becomes a policy-enforced object.
Governance across classification tiers	Federated architecture. Collibra governs the catalog. Data X-Ray scans on-prem, hybrid, and cloud. Intelligence flows without files leaving the environment. Sensitive assets flagged in cross-tier migrations.
Varonis replacement readiness	Deploys into existing RHEL infrastructure. Operational in hours. Addresses documented Varonis gaps: redaction (scored 1 out of 5, Forrester Wave Q1 2025), governance integration (none), continued on-prem support.
Security certifications	STIG-compliant. ATO artifacts provided. Letter of Attestation from independent auditor. Requirements Traceability Matrix for accelerated ATO.



Deployment Risk: BigID in This Environment

BigID is built for commercial environments. For deployments with strict air-gap, on-prem, and unstructured data requirements, three gaps present material deployment risk.

Air-gapped deployment. BigID is cloud-native. Its own documentation describes on-prem local scanner setup as historically complex and resource-intensive, with admins spending days or even weeks getting a single scanner operational. (BigID blog, October 2025.) FedRAMP authorization has not been achieved as of February 2026. BigID is pursuing it through a managed cloud partner, Knox Systems, with no native air-gapped authorization path. (PR Newswire, December 2025.)

Unstructured data classification. A Deloitte consultant with 2+ years of production experience on BigID noted in a verified review that the platform should focus more on the unstructured data part, and that seamless unstructured scanning comparable to its structured capability is not yet there. (PeerSpot, 2025.) In an environment where the majority of sensitive mission data lives in unstructured files, this is a foundational gap.

Architectural Lock-In: A Structural Risk in a Defined Stack

BigID's model is consolidation: intelligence stays inside its proprietary ecosystem. Gartner noted in September 2025 that DSPM data catalogs remain proprietary and do not support open standards, leading to **significant vendor** lock-in. In a DoW environment where the security stack is already defined, a platform that enriches existing tools without replacing them is the lower-risk architectural choice.

Next Step

The architecture is validated. The certifications are in place. The stack is deployable into existing infrastructure today.

The next step is a direct technical session to map DoW's AOA criteria to demonstrate capability. Request that session here: <https://www.ohalo.co/demo>

All BigID competitive claims are sourced to publicly available materials: BigID's own published documentation (bigid.com), Gartner Peer Insights, PeerSpot, Forrester Wave Q1 2025, and PR Newswire. Full source citations available on request.



Schedule a Conversation

Scan the QR code or visit us online

ohalo |  Data X-Ray

www.ohalo.co