A Metadata-Driven Secure Data-Sharing Model for Controlled Access to Proprietary Information

Author: Eddie A. Cabán Ferrer
Advisor: Jeffrey Duffany, Ph.D.
Master in Computer Science
Graduate Project EXPO, May 2026

Abstract

A metadata-driven secure data-sharing model is proposed for controlling access to proprietary information exchanged between internal enterprise sources and external collaborators. The framework was designed to show that secure sharing can be governed through identity-aware authorization, tags-based policy enforcement, protected file storage, and persistent auditing rather than through static permissions or vendor-specific platforms. The implemented system integrates an identity provider, a policy evaluation engine, a relational database for metadata and audit records, object storage for protected files, and an application layer that coordinates token validation, authorization queries, storage access, and request measurement. Evaluation was conducted in three stages: a controlled baseline, a moderate workload stage, and a larger workload stage. Across these stages, the framework maintained correct access-control behavior, prevented unauthorized proprietary downloads, and preserved complete audit capture for protected actions. The results support the feasibility of a vendor-independent middleware approach for deterministic, auditable, policy-directed control of sensitive document sharing.

Introduction

Organizations increasingly depend on digital systems to create, manage, and exchange proprietary technical and business information. The literature on enterprise information management shows that proprietary information control must be understood as part of a broader organizational process that spans engineering, operational, and collaborative environments rather than as a simple file-storage problem [1]. In practice, that information rarely remains inside a single application or department. Instead, it moves across engineering, operational, and collaborative environments where different stakeholders require different levels of visibility and control. This fluid movement suggests that secure sharing must be treated as a multifaceted information governance challenge rather than a simple file transfer problem.

Background

Current discourse in the field emphasizes a shift toward identity aware authorization and metadata-driven control, where proprietary files are classified at the point of intake and managed through protected object storage. One of the motivators for this shift is the mitigation of insider threats and accidental data leakage, which are often exacerbated by static permission models that fail to adapt to the data's lifecycle or context. Policy-enforced gateways and deterministic authorization provide the foundation for modeling the rigorous governance found in industrial Product Lifecycle Management (PLM) and Enterprise Resource Planning (ERP) systems. Furthermore, a robust emphasis on auditability ensures that every interaction with a proprietary file leaves a chronological, tamper-evident trail. This traceability is essential for forensic analysis and compliance verification. As a result, this approach places the management of sensitive data at the critical intersection of secure system design, auditable enforcement, and strategic enterprise oversight. In this setting, identity-aware authorization, configuration-driven enforcement, and structured auditability emerge not merely as technical features, but as core governance mechanisms for controlling sensitive information across organizational boundaries [2][3][5].

Problem

Businesses frequently need to share proprietary technical and corporate information with internal users, suppliers, contractors, and partner entities. This sharing often occurs through fragmented systems that do not combine authentication, authorization, object storage, and auditability in a consistent manner. Consequently, organizations face recurring risks including unauthorized disclosure, over-permissioned access, inconsistent policy enforcement, limited visibility into who accessed what, and difficulty proving compliance or accountability after an incident or review. This project responds to such gap by modeling a focused set of secure-sharing functions using an open-source software stack. In this design, authenticated users receive identity-bearing tokens, file objects are classified and associated with required access attributes, metadata and audit evidence are stored authoritatively, policy decisions are evaluated deterministically before release, and all successful and denied actions are recorded for later analysis. The objective is not to replicate a full commercial enterprise platform, but to model key subsets of the governance and control functions in PLM, ERP, and collaboration-oriented environments. The project is grounded in the idea that secure sharing must be treated as an information-governance problem rather than a simple file-transfer problem.

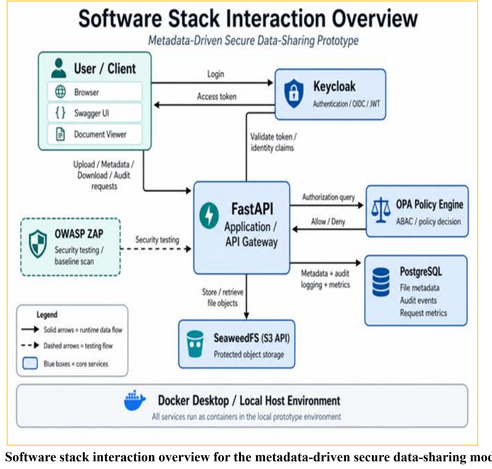
Methodology

The system was deployed as a hybrid local environment on a single Windows laptop. Containerized services were orchestrated using Docker Desktop, while object storage ran as a separate host-level process to ensure persistent local data retention without relying on heavy enterprise infrastructure. the architecture integrates FastAPI, Keycloak, Open Policy Agent (OPA), PostgreSQL, SeaweedFS, and OWASP ZAP to collaboratively manage application orchestration, identity provision, policy enforcement, relational storage, object storage, and baseline security validation.

Component	Role	Why It Matters
FastAPI	Application gateway and API layer	Implements upload, metadata, download, audit, and identity-check endpoints.
Keycloak	Identity provider	Issues user tokens and centralizes user and role management.
OPA	Policy engine	Evaluates decisional allow/deny decisions from token and file metadata.
PostgreSQL	Metadata and audit store	Stores file records, request metrics, and audit events.
SeaweedFS	Object storage	Persists uploaded files behind an S3-compatible interface.
OWASP ZAP	Security testing tool	Provides passive scanning and baseline web-surface validation.

Software stack components and their roles in the framework.

The operational mechanics of this architecture are illustrated by the framework workflow. The interaction begins when a user authenticates through Keycloak and receives a cryptographically signed JSON Web Token (JWT) containing trusted identity claims, such as their role, department, and clearance level. When the user submits an API request, the FastAPI gateway intercepts the call, validates the JWT signature using Keycloak's public key, and extracts the identity claims. Concurrently, FastAPI retrieves the requested file's informational tags and required access attributes from PostgreSQL. FastAPI then structures this combined identity and metadata context and submits it as an authorization query to OPA. OPA evaluates these inputs against the predefined Attribute-Based Access Control (ABAC) policies. If OPA issues an "allow" decision, FastAPI brokers access to the physical file residing in SeaweedFS and streams it to the user. If the decision is "deny," the request is rejected. Regardless of the outcome, all policy-relevant file actions are persistently recorded in the PostgreSQL audit log.



Software stack interaction overview for the metadata-driven secure data-sharing model.

To quantify system behavior, the application was instrumented with request-level timing middleware capturing metrics like latency and authorization decisions. The evaluation of this practice was conducted in three progressive stages. The Baseline stage utilized a controlled matrix of 25 test cases across 4 seeded files to verify the exact correctness of the access logic. Tier 1 introduced a moderate workload featuring 10 users and 100 documents to confirm stability under automated multi-user operations. Finally, Tier 2 expanded the scale to 20 users, 500 documents, and 1,520 client-side operations to assess the model's endurance and scalability under a significantly larger workload while preserving the same authorization logic.

Results and Discussion

The evaluation of the tags-driven data-sharing model progressed through three distinct stages: a controlled Baseline, a Tier 1 moderate workload, and a Tier 2 larger workload. The results demonstrated that the framework maintained strict, deterministic access control and flawless auditability across all operational scales without requiring changes to the underlying access-control model. In the initial Baseline stage, the system was subjected to a correctness matrix of 25 controlled test cases.

Category	Cases	Observed Result	Outcome
Seed uploads	4	4/4 succeeded	Pass
Authorized proprietary access	8	All returned 200 and allowed	Pass
Unauthorized proprietary access	5	All returned 403 and denied	Pass
Public-file access	4	All returned 200 and allowed	Pass
Audit endpoint	1	Returned 200	Pass
No-token / expired-token cases	3	All returned 401	Pass
Role-mutation behavior	4	Matched uploader-override design	Pass

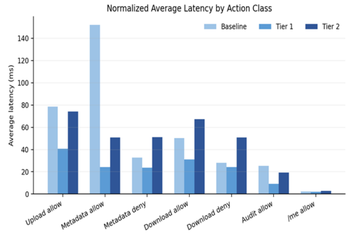
Baseline correctness summary across controlled case categories.

The model achieved a 100.00% policy accuracy rate, successfully allowing authorized interactions while consistently blocking unprivileged access to proprietary files . The subsequent evaluation stages confirmed the framework's reliability under pressure. Tier 1 processed 100 documents and 331 operations, while Tier 2 scaled significantly to 500 documents and 1,520 operations.

Metric	Baseline	Tier 1	Tier 2
Directed / client-side cases	25	331	1,520
Documents uploaded	4	100	500
Client pass rate	100.00%	99.70%	100.00%
Unauthorized download success rate	0.00%	0.00%	0.00%
False allow rate	0.00%	0.00%	0.00%
False deny rate	0.00%	0.00%	0.00%
Audit completeness	100.00%	100.00%	100.00%
Audit rows recorded	25	300	1,460
Server-side requests recorded	111	332	1,520
Authentication failures (401)	3	2	0
Notes	Directed interactive baseline	One /me mismatch only	Normalized Tier 2 rerun

Comparative outcome summary for Baseline, Tier 1, and Tier 2.

The comparative outcomes captures reveal the project's most significant security achievement: the system maintained a 0.00% unauthorized download success rate, a 0.00% false allow rate, and a 0.00% false deny rate across all tiers . The system actively prevented unauthorized retrieval while maintaining 100.00% audit completeness, perfectly capturing all 1,460 audit-eligible actions during the heavy Tier 2 run.



Normalized average latency by action class across Baseline, Tier 1, and Tier 2.

While the system experienced a measurable increase in latency as the workload scaled, these increases remained bounded and stable. Crucially, this scaling cost did not degrade the correctness of the authorization outcomes, proving the local feasibility of the architecture.

Conclusions

This project demonstrated a feasible metadata-driven secure-sharing framework using a working middleware model and an open-source, vendor-independent software stack. Across all evaluation tiers, the system maintained correct access-control behavior, prevented unauthorized proprietary downloads, and preserved complete audit capture for protected actions. These findings prove that identity-aware, tags-driven, and auditable authorization can regulate proprietary document sharing instead of static permissions or platform-specific controls. While designed for enterprise-style information governance, this model is equally relevant to smaller organizations managing sensitive client or regulated records requiring traceable, policy-based release. The project contributes both a local proof-of-concept and a general framework for environments where accountability and explainable access decisions are essential. Utilizing containerized services and host-level object storage creates a modular, portable stack that can be managed locally or migrated to larger environments. Although not a test of enterprise-scale readiness, this implementation shows that identity management, policy enforcement, and secured storage can be successfully integrated without a commercial platform. Such an approach is specifically applicable to paperless engineering, manufacturing, and other record-driven workflows where sensitive documents must be governed over time while reducing the risks associated with fragmented file-handling practices.

Future Work

Future work should expand the framework to distributed and repeatable operating environments. The results demonstrate local feasibility and provide a solid foundation for studying authentication, policy enforcement, meta-information retrieval, object storage access, and audit logging across multiple virtual machines or cloud instances under more realistic deployment conditions. To reliably recreate assessment phases, upcoming research might improve reproducibility through automatic environment reset, seeded dataset generation, programmed workload replay, and structured figure production. Richer audit analytics, expressive attribute-based policy conditions, secure document preview, viewer-layer controls, and integration-friendly APIs can enhance the framework. These strategies establish a reliable foundation for a metadata-driven secure-sharing framework that is mature, repeatable, scalable, and extendable.

Acknowledgements

Special thanks to the procurement personnel that contributed to this work, as well as Prof. Jeffrey Duffany, PhD for all the support and guidance through this project.

References

1. S. D. Murthy, "The Strategic Architecture of Modern Manufacturing: PLM and ERP as Complementary Digital Systems," Journal of Information Systems Engineering and Management, vol. 11, no. 1s, 2026.
2. A. Chiquito, U. Bodin, and O. Schelen, "Automated Management of Attribute-Based Policies for Access Control using Tag-Matching," 2023.
3. V. Biagi and A. Russo, "Data Model Design to Support Data-Driven IT Governance Implementation," Technologies, vol. 10, no. 106, 2022, doi: 10.3390/technologies10050106.
4. S. D., Minavathi, and M. S. D., "Towards Resilient PLM Architectures: Cyber Threats, Security Mechanisms, and Industrial Implications," IITM Journal of Management and IT, vol. 16, no. 2, Dec. 2025.
5. S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture," National Institute of Standards and Technology (NIST), Gaithersburg, MD, Special Publication (SP) 800-207, 2020. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-207
6. A. Teymurova, "Metadata Management in Data Governance: A Comprehensive Scientific Analysis," *Azerbaijan State University of Economics (UNEC)*, Baku, Azerbaijan, 2026. [Online]. Available: teymurova.arzu.rovshan.2023@unec.edu.az
7. A. O. Adebayo, "The SAIS-GRC Framework: Engineering Trust and Secure, Agile Systems for Proactive AI Governance and Compliance," *Iconic Research and Engineering Journals*, vol. 8, no. 5, pp. 1475-1480, Nov. 2024. doi: 10.64388/IREV7I5-1713349.