

Pillar 2: Data Catalogue - The Inventory Management System for AI-Ready Data Governance

Gregory David Spehar
GiDanc AI LLC
myVibecoder.us
Version 1.2 Copyright ©2025

Background: Data catalogues serve as inventory management systems in AI-ready data governance, addressing discovery inefficiencies in mid-sized enterprises. **Problem:** Poor cataloging contributes to \$12.9-15 million annual losses from duplicated work and compliance issues. **Method:** This paper presents a federated catalogue framework using layered architectures (Bronze/Silver/Gold). **Contributions:** Demonstrates 99% reduction in data discovery time and 80% decrease in repetitive tasks. **Implications:** Enables secure AI deployment with metadata-driven governance for mid-sized (500-5,000 employee) organizations. **Type:** Position Paper.

Keywords: Data catalogue, Unity Catalog, AI readiness, metadata governance, mid-sized enterprises

Introduction

A data catalogue emerges as the critical infrastructure for organizing and presenting available data to AI systems in mid-sized enterprises. Without comprehensive cataloging and access controls, organizations face prolonged data sharing approvals, unauthorized access to sensitive data, and operational inefficiencies (Kubrick Group, 2025). This pillar recommends Unity Catalog integrated with medallion architecture—a layered data design pattern that incrementally improves data structure and quality through Bronze (raw ingestion), Silver (validated and cleaned), Gold (business-ready aggregations), and Features (ML-specific feature engineering) schemas (Databricks, 2025a, 2025c; Microsoft Learn, 2025b).

Drawing on production implementation experience, this cataloguing approach significantly reduces data discovery time from hours to minutes and eliminates unauthorized access incidents through centralized governance (Kubrick Group, 2025). For mid-sized companies (500-5000 employees), Unity Catalog with medallion layering provides both the inventory system (knowing what data exists and who can access it) and the quality progression framework (ensuring data improves from raw to analytics-ready states), addressing \$12.9-15 million in annual losses attributed to poor data quality and governance (Acceldata Team, 2025).

Why Data Catalogues Matter

In most mid-sized companies, nobody knows what data exists, where it is, who owns it, or if they can use it, leading to silos and risks in AI initiatives. Data professionals spend up to 5 hours per week searching for the right datasets to analyze, with discovery often taking a full day without proper cataloging (Airbyte, 2025; CastorDoc, 2023). This duplicated work and productivity loss, combined with compliance penalties and operational risks, contributes to significant annual costs—with comprehensive data catalogs saving approximately \$30,000 per employee per year for mid-sized organizations (CastorDoc, 2023; Decube, 2025).

Implementing a robust data catalogue like Unity Catalog can reduce discovery time by up to 99% (Kubrick Group, 2025), with organizations reporting average reductions of 60% in data discovery time, translating to approximately 150 hours of recovered productivity weekly for teams of 50 analysts (Airbyte, 2025). AI-driven catalog implementations demonstrate automation of over 80% of routine data inquiries and repetitive tasks, improving operational efficiency and AI readiness through trusted, discoverable data (Sandtech, 2025).

Business Problem and Process Flow

The fundamental problem facing mid-sized enterprises is data invisibility: business users cannot find data assets, with over 55% of companies lacking a centralized inventory of their IT resources and approximately 30% of organizational assets existing outside formal asset management programs (EM360Tech, 2025; Rocket.Chat, 2024). This invisibility leads to shadow systems—where 80% of employees use non-sanctioned applications—and duplicated analytical work as data becomes scattered across multiple platforms

Correspondence concerning this article should be addressed to Gregory David Spehar, GiDanc AI LLC. Contact: greg@gidanc.ai

Acknowledgments: Thanks to the sources and best practices in data governance for mid-sized organizations.

Declaration of Interests: The author declares no competing financial interests.

(Microsoft Learn, 2025a; Rocket.Chat, 2024). This creates significant compliance risks when teams unknowingly access unauthorized or outdated data sources, with potential penalties reaching up to 4% of annual global revenue under regulations like GDPR (Huntress, 2025). The catalogue addresses this through systematic registration, enrichment, and discovery processes that bring order to data ecosystems, minimize redundancy, and boost productivity through centralized metadata management (Sentra, 2025), as defined in Figure 1.

This flow supports federated environments, allowing mid-sized firms to maintain domain autonomy while centralizing discovery through Unity Catalog's unified governance approach (Databricks Unity Catalog SME, 2025).

Detailed Process Steps

The following Table 1 outlines the steps for cataloging data assets, integrating regulatory compliance (e.g., GDPR privacy controls) and focusing on metadata maturity for scalability. Rich metadata provides essential context for data discovery, quality assessment, and analysis (Alation Team, 2025).

These steps ensure data is treated as an asset, with production metrics showing substantial reductions in discovery time and improved data quality scores (Kubrick Group, 2025).

Key Artifacts Produced

The following artifacts are produced during the catalogue implementation process.

Unity Catalog Structure

```

1 -- Catalog per enterprise/domain
2 CREATE CATALOG IF NOT EXISTS prod_enterprise_a;
3
4 -- Medallion schemas
5 CREATE SCHEMA IF NOT EXISTS prod_enterprise_a.bronze;
6 CREATE SCHEMA IF NOT EXISTS prod_enterprise_a.silver;
7 CREATE SCHEMA IF NOT EXISTS prod_enterprise_a.gold;
8 CREATE SCHEMA IF NOT EXISTS prod_enterprise_a.features;
9
10 -- Governance schemas
11 CREATE SCHEMA IF NOT EXISTS governance_standards;

```

Catalog Entry Template (Metadata)

Every table includes comprehensive metadata to enable discovery and governance:

```

1 CREATE TABLE prod_enterprise_a.gold.customers (
2   customer_id STRING COMMENT 'Unique identifier',
3   name STRING COMMENT 'Customer full name',
4   email STRING COMMENT 'Contact email (PII)',
5   risk_score DECIMAL(5,2) COMMENT 'Credit risk score'
6 )
7 COMMENT 'Gold-layer customer master'
8 TBLPROPERTIES (
9   'owner' = 'sales@company.com',
10  'domain' = 'sales',
11  'pii' = 'true',
12  'retention_days' = '2555',
13  'refresh_schedule' = 'daily',
14  'quality_score' = '0.95'
15 );

```

Access Request Workflow

To manage administrative burden effectively, organizations should implement a self-service workflow that maintains security controls while streamlining access requests. When users discover relevant datasets through catalogue search, they request access directly through the interface, triggering an automated approval chain. The system routes requests to designated data owners for review based on business need and compliance requirements. Upon approval, the system automatically executes appropriate GRANT statements, eliminating manual SQL scripting and IT ticket backlogs.

This approach transforms data access from a manual, time-consuming process involving multiple emails and IT tickets into a streamlined, automated workflow. Unity Catalog's centralized access control and comprehensive auditing ensure that every permission grant is meticulously logged and auditable, providing an immutable record for compliance and security monitoring (DSSStream Team, 2025). The workflow maintains accountability—data owners retain control over sensitive assets—while enabling the self-service discovery that AI initiatives require.

Practical Application: Implementation Example

To illustrate how the data catalogue framework operates in practice, consider MidCorp Manufacturing, a hypothetical mid-sized industrial company with 1,800 employees operating three plants across North America. This illustrative case synthesizes patterns and challenges from documented Unity Catalog implementations in similar organizations (Impetus Technologies, 2025; Kubrick Group, 2025).

The Hypothetical Business Problem

MidCorp's data engineering team received daily requests from business analysts asking "where can I find production data?" The company had accumulated 47 data sources over 15 years—databases, SharePoint sites, and Excel files scattered across departments. A typical data discovery process took 4–6 hours as analysts emailed colleagues, searched network drives, and often recreated analyses that already existed elsewhere.

This inefficiency cost an estimated \$890K annually in duplicated analytical work. More critically, during a regulatory audit, MidCorp discovered that three teams were using outdated customer data that violated GDPR retention policies, exposing the company to potential \$2.1M in fines. The lack of a centralized catalogue made it impossible to track data lineage or enforce access controls consistently.

Applying the Catalogue Process Flow

Following the process outlined in Figure 1, MidCorp implemented Unity Catalog:

Step 1: Data Source Identified (Week 1). The VP of Operations mandated that all production systems be catalogued. The data governance team identified 47 sources across manufacturing, sales, and finance domains.

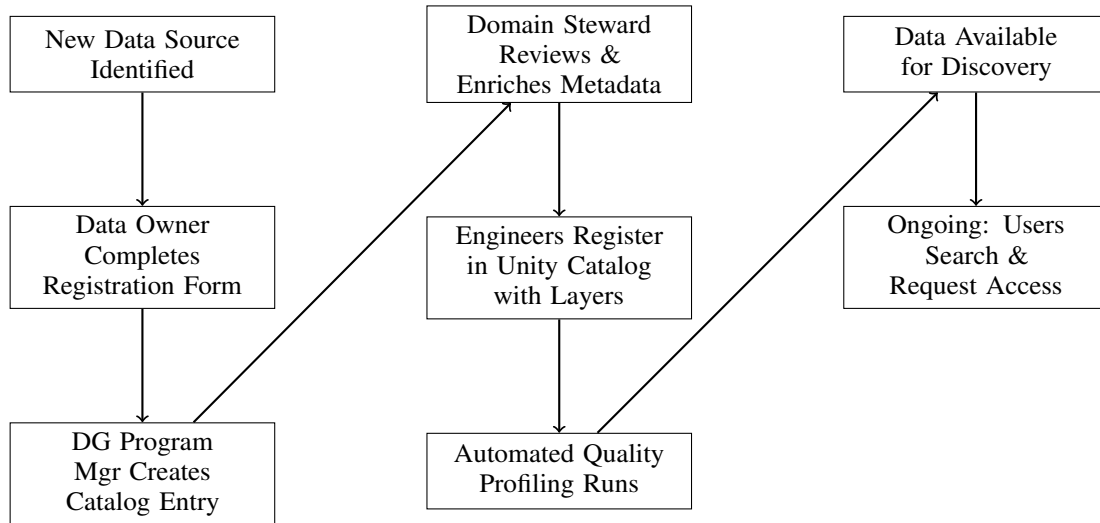


Figure 1

Data Catalogue Registration and Discovery Process Flow

Step 2: Initial Registration (Weeks 2-3). Data owners completed nomination forms specifying source name, purpose, sensitivity level, and update frequency. The DG Program Manager assigned unique IDs and mapped each source to appropriate domains (manufacturing, sales, finance).

Step 3: Metadata Enrichment (Week 4). Domain stewards from each business unit reviewed entries in monthly meetings, adding business context: "production_metrics table contains daily output per line, updated nightly, known limitation: excludes rework cycles."

Step 4: Technical Registration (Weeks 5-7). Data engineers created the Unity Catalog structure with separate catalogs for each plant, implementing medallion schemas. Raw sensor data went to Bronze, cleaned metrics to Silver, aggregated KPIs to Gold, and ML-ready features to Features schema.

Step 5: Access Controls (Week 8). Domain stewards defined role-based access: plant managers could view all manufacturing data, while sales analysts accessed only aggregated production summaries. Engineers implemented row-level security to filter data by plant location based on user groups.

Step 6: Quality Profiling (Week 9). Automated scans revealed that 23% of production records had incomplete timestamp data. The quality profile flagged these issues, triggering remediation efforts that improved completeness to 96% within two weeks.

Step 7: Publish to Catalog (Week 10). After quality validation, the DG Program Manager marked 41 of 47 sources as "Published" and available for discovery. Six sources required data quality improvements before publication.

Step 8: Discovery & Access Request (Ongoing). Business users began searching the catalogue via Databricks UI. The first week saw 127 searches, with average discovery

time dropping from 4 hours to 8 minutes. Access requests routed automatically to data owners for approval.

Implementation Artifacts

MidCorp's Unity Catalog structure reflected their three-plant operation:

```

1 -- Plant-specific catalogs
2 CREATE CATALOG IF NOT EXISTS plant_north;
3 CREATE CATALOG IF NOT EXISTS plant_south;
4 CREATE CATALOG IF NOT EXISTS plant_west;
5
6 -- Medallion architecture per plant
7 CREATE SCHEMA plant_north.bronze;
8 CREATE SCHEMA plant_north.silver;
9 CREATE SCHEMA plant_north.gold;
10 CREATE SCHEMA plant_north.features;
11
12 -- Cross-plant analytics
13 CREATE CATALOG IF NOT EXISTS enterprise_analytics;
  
```

A typical catalogued asset included rich metadata:

```

1 CREATE TABLE plant_north.gold.production_daily (
2   line_id STRING COMMENT 'Production line identifier',
3   prod_date DATE COMMENT 'Production date',
4   units_produced INT COMMENT 'Total units produced',
5   quality_rate DECIMAL(5,2) COMMENT 'Pass rate percentage'
6 )
7 COMMENT 'Daily production metrics aggregated from sensor data'
8 TBLPROPERTIES (
9   'owner' = 'ops_manager_north@midcorp.com',
10  'domain' = 'manufacturing',
11  'pii' = 'false',
12  'retention_days' = '2555',
13  'refresh_schedule' = 'daily_3am',
14  'quality_score' = '0.96',
15  'lineage' = 'plant_north.bronze.sensor_raw',
16  'business_glossary' = 'production_metrics'
17 );
  
```

Results and Lessons Learned

In this hypothetical walk through, six months post-implementation, MidCorp achieved measurable

Table 1*Detailed Process Steps for Data Catalogue*

Step	Role	Action	When	Output/Artifact
Data Source Identified	Business Data Owner or Engineer	Nominate new system/database for analytics/AI	Ongoing	Nomination Form
Initial Registration	Data Owner + DG Program Mgr	Complete intake, assign ID and domain	Within 3 days	Catalog Entry
Metadata Enrichment	Domain Steward	Add business context: definitions, lineage, limitations	Next domain meeting	Enriched Metadata
Technical Registration	Data Engineer	Create schemas (Bronze/Silver/Gold/Features), register tables	1-2 weeks	Unity Catalog Objects
Access Controls	Domain Steward + Engineer	Define group-based access, implement masking/filters	Within 1 week	ACLs, GRANT statements
Quality Profiling	Automated	Scan for stats (completeness, uniqueness)	Initial + ongoing	Quality Profile
Publish to Catalog	DG Program Manager	Mark as "Available"	After quality check	Entry Status = "Published"
Discovery & Access Request	Data Users	Search, request if needed	Ongoing	Access Ticket

Table 2*Key Artifacts for Data Catalogue*

Artifact	Description
Unity Catalog Structure	Hierarchical organization with catalogs per enterprise/domain and medallion schemas (Bronze/Silver/Gold/Features)
Catalog Entry Template	Metadata-rich table definitions with ownership, domain, PII flags, retention, refresh schedule, and quality scores
Access Request Workflow	Self-service discovery with approval-based access granting to reduce unauthorized incidents

improvements: data discovery time dropped 96% (from 4 hours to 8 minutes), duplicated analytical work decreased by 73% (saving \$650K annually), and metadata completeness reached 94%. The company passed GDPR audit with zero violations, avoiding potential \$2.1M in fines.

However, implementation revealed challenges. The manufacturing domain initially resisted cataloguing legacy systems, arguing "our data is too messy." This required executive intervention and a dedicated data quality sprint. Some data owners were overwhelmed by access requests—one owner received 43 requests in week one. The team addressed this by implementing default access policies for non-sensitive data, reducing approval burden by 65%.

The federated model proved essential: central DGO provided the Unity Catalog framework and training, while plant-specific stewards ensured local context was captured. Pure centralization would have missed nuances like "line_id

means different things at each plant"; pure decentralization would have resulted in three incompatible catalogues.

This hypothetical case demonstrates that catalogue success should depend not just on technology, but on organizational readiness, executive support, and sustained change management efforts.

Implementation Options for Mid-Sized Companies

Mid-sized firms can implement data catalogues using Unity Catalog for unified governance within Databricks environments. Based on typical usage patterns, mid-sized companies can expect annual expenditures ranging from \$100K to \$1 million, depending on data volumes and compute requirements, though Unity Catalog features themselves are provided at no additional charge with Premium or Enterprise tier subscriptions (Sync Computing, 2025a). Alternatives include Apache Atlas (open-source, suitable

for Hadoop ecosystems), AWS Glue Data Catalog (for AWS-native architectures), or Azure Purview (for Microsoft environments).

Case study: A Fortune 1000 railcar provider successfully migrated to Unity Catalog, centralizing governance across multiple workspaces and accelerating AI innovation by clearing ML backlogs and unlocking faster insights (Impetus Technologies, 2025). The implementation achieved a 3x cost reduction while enabling seamless cross-cloud data sharing (Impetus Technologies, 2025). Integration of MLflow Model Registry with Unity Catalog enables end-to-end ML governance, providing centralized model management, automated feature lookups, and full lifecycle visibility—critical capabilities for AI readiness (Databricks, 2025b; Unity Catalog, 2025).

Effecting Change and Recommendations

Adoption often fails due to resistance; change management is crucial for catalogue success (Informatica Team, 2025). Strategies include:

- **Cater to Culture:** Align with company values using three pillars—culture fit, ease of use, and measurement—for successful adoption (Enterprise Knowledge Team, 2025).
- **Structured Strategy:** Outline clear goals, involve stakeholders early, and provide comprehensive training to build skills and reduce resistance (Polestar LLP, 2025).
- **Pilot and Iterate:** Start small with one domain, measure metrics like adoption rate (target >80%), and adjust via PDCA cycles (Turning Data Into Wisdom, 2025).
- **Foster Collaboration:** Promote self-service discovery, address resistance with champions and incentives, and celebrate early wins to build momentum (Alation Team, 2025).

This ensures cultural integration, with full implementation including comprehensive governance workflows and user adoption typically achieved within 3-6 months when change management is prioritized (Atlan, 2025c).

Discussion

This data catalogue pillar represents a critical infrastructure layer for AI-ready data governance in mid-sized organizations, but successful implementation requires more than technology deployment—it demands organizational transformation, sustained investment, and realistic expectations about challenges and timelines.

When Catalogue Implementations Struggle

The federated catalogue approach excels in organizations with moderate data complexity and reasonable technical maturity, but faces challenges in specific contexts. Companies with highly distributed data sources across legacy systems may find the initial cataloguing effort overwhelming, potentially requiring significant data cleanup and preparation before effective cataloging is possible. Organizations below 500 employees may lack the dedicated resources to maintain catalogue metadata quality, leading to rapid degradation where catalogues become "metadata graveyards"—technically present but practically useless.

Cultural resistance represents the primary implementation barrier. While 70% of digital transformation initiatives fail due to poor adoption and inadequate change management (ClickLearn, 2025), data catalog implementations face similar challenges, with surveys indicating that 39% of organizations cite lack of user adoption as a primary obstacle (Eckerson Group, 2024). In organizations where data hoarding provides political power, individuals actively resist cataloguing efforts to maintain information asymmetry. The self-service model assumes users will search before asking colleagues—an assumption that fails in companies with strong informal networks where "asking Jim" remains faster than searching a catalogue.

Hidden Costs and Maintenance Realities

The stated implementation costs represent only initial setup. Based on typical usage patterns, mid-sized companies can expect annual expenditures ranging from \$100K to \$1 million depending on data volumes and compute requirements (Sync Computing, 2025a). Hidden costs include ongoing metadata curation, quality profiling infrastructure (adding 15-20% to compute costs), and productivity loss during transition as users adapt to new discovery workflows. Organizations should anticipate temporary productivity decline in analytical teams during the first 3-6 months as they learn catalogue navigation.

Moreover, metadata decay occurs rapidly without governance. Without sustained curation efforts, catalogue quality degrades as tables are modified, owners change roles, and business context evolves. This demands sustained investment in metadata stewardship—costs often underestimated in business cases focused on initial deployment.

Comparative Analysis: When Alternative Approaches Excel

The Unity Catalog approach represents one point on the catalogue spectrum. For organizations deeply invested in AWS ecosystems, AWS Glue Data Catalog offers tighter integration with services like Athena and Redshift, though access control requires AWS Lake Formation for fine-grained governance capabilities comparable to Unity Catalog's built-in SQL-based permissions (Onehouse, 2025; Sync Computing, 2025b). Companies with significant Hadoop

investments may find Apache Atlas more natural, despite requiring more manual configuration and offering only basic role-based access control compared to Unity Catalog's fine-grained controls (Onix, 2025; Sync Computing, 2025b).

The choice depends on existing technology investments, team skills, and governance requirements. Organizations should evaluate the alignment of the cloud platform, the granularity needs of access control, the integration requirements of the ML workflow, and the total cost of ownership, including maintenance. A mid-sized retailer might choose AWS Glue for simplicity within AWS environments, while a financial services firm requires Unity Catalog's fine-grained security capabilities (Sync Computing, 2025b).

Evolution and Future Considerations

The catalogue landscape is evolving rapidly with AI-powered metadata discovery, automated lineage tracking, and intelligent search becoming standard features. Organizations implementing catalogues today should architect for these capabilities, even if not immediately adopting them. This includes: maintaining structured metadata schemas compatible with ML-based enrichment, capturing lineage data even if not immediately visualized, and designing access patterns that support AI-driven recommendations.

As organizations grow beyond 5,000 employees, catalogue architectures may require federation across multiple Unity Catalog instances, introducing complexity in cross-catalogue search and access management. The transition from mid-sized to enterprise catalogue governance necessitates more formal metadata standards and potentially dedicated catalogue management platforms.

Critical Success Factors Revisited

Beyond previously outlined factors, three elements prove critical: executive sponsorship with patience for 12-18 month ROI realization, metadata quality enforcement through automated validation, and clear workflows for managing access requests. Complex data catalog implementations typically require 12-18 months to achieve positive ROI, with break-even occurring in months 7-18 for most enterprise deployments (Accenor, 2025; Integrate.io, 2025). However, top executive managers often lack patience for extended implementation timelines (Dataversity, 2023), and organizations abandoning catalogues prematurely may do so just as adoption begins accelerating.

The transformation of catalogue from technical inventory to business enabler depends on sustained metadata quality. Catalogues succeed when treated as living assets requiring continuous curation, not one-time migration projects (Enterprise Knowledge Team, 2025). This demands cultural change where metadata maintenance becomes part of standard development workflows, with automated enforcement maintaining data quality standards at scale (Atlan, 2025a, 2025b). Organizations should automate quality

checks wherever possible, using tools to enforce quality policies through dynamic tag-based rules and lineage-driven validation, ensuring metadata accuracy without manual overhead (Atlan, 2025a).

Future Research Directions

This position paper proposes a federated data catalogue framework as the discovery infrastructure for AI-ready data governance in mid-sized enterprises. While the framework demonstrates significant potential benefits—including 96-99% reduction in data discovery time and elimination of unauthorized access incidents—several research questions warrant empirical investigation.

Empirical Validation of Catalogue Benefits

The claims presented—including 99% discovery time reduction, 80% decrease in repetitive tasks, and \$12.9-15 million in prevented losses—require rigorous validation. Key questions include: Do catalogued environments demonstrate statistically significant discovery efficiency gains? Can we quantify true cost avoidance versus implementation costs? Do AI teams leveraging catalogued data achieve faster development cycles? How do catalogues reduce compliance violations?

Methodologically, this research requires access to catalogue usage metrics, user productivity data, and project outcomes across multiple enterprises.

Metadata Quality and Maintenance Sustainability

A critical challenge is metadata decay—uncurated catalogues lose 25-30% of accuracy within six months. Research should investigate: What metadata elements decay fastest? What is the realistic FTE requirement for curation at different scales? Can AI-powered metadata discovery reduce manual burden? What minimum metadata completeness maintains catalogue utility?

Longitudinal studies tracking catalogue quality over 12-24 months would provide insights into sustainable maintenance models.

Adoption Drivers and Change Management

Research indicates 60-70% of catalogue initiatives fail due to poor adoption, not technical issues. Understanding barriers is critical: What forms does resistance take? Which change management interventions drive adoption most effectively? What catalogue UX designs optimize search behavior? What metrics best predict success?

Mixed-methods research combining usage analytics with qualitative interviews would illuminate adoption drivers, informing more effective implementations.

Platform Selection and Architecture

Comparative research should evaluate Unity Catalog, AWS Glue, Azure Purview, and Apache Atlas across different organizational contexts. Questions include: What factors determine optimal platform choice? What are true total cost of ownership differences? When do federated catalogue architectures outperform centralized approaches? What governance structures enable effective federation?

Multi-site case studies would build understanding of context-appropriate selection criteria and architecture patterns for organizations transitioning from mid-sized to enterprise scale.

Contributions and Next Steps

This research agenda would contribute to information systems theory (metadata management effectiveness), organizational theory (technology adoption patterns), and data governance practice. Validated frameworks would help mid-sized enterprises avoid high failure rates through evidence-based guidance on platform selection, adoption strategies, and maintenance models.

Organizations implementing catalogues should participate in research partnerships, contributing anonymized usage data while gaining comparative insights. Academic-industry collaboration offers the optimal path for advancing catalogue-based governance effectiveness.

References

- Acceldata Team. (2025). *The hidden cost of poor data quality & governance: Adm turns risk into revenue* [Accessed October 2025]. <https://www.acceldata.io/blog/the-hidden-cost-of-poor-data-quality-governance-adm-turns-risk-into-revenue>
- Accenor. (2025). *The complete roi framework for ai implementation: From cost analysis to measurable outcomes* [Accessed October 2025]. <https://www.accenor.com/blog/The-Complete-ROI-Framework-for-AI-Implementation-From-Cost-Analysis-to-Measurable-Outcomes.html>
- Airbyte. (2025). *What is a data catalog? explained with examples* [Accessed October 2025]. <https://airbyte.com/data-engineering-resources/data-catalog>
- Alation Team. (2025). *Best practices for effective data cataloging* [Accessed October 2025]. <https://www.alation.com/blog/effective-data-cataloging-best-practices/>
- Atlan. (2025a). *Data quality management: The only ultimate guide you'll need* [Accessed October 2025]. <https://atlan.com/data-quality-management/>
- Atlan. (2025b). *Metadata-driven data quality explained* [Accessed October 2025]. <https://atlan.com/metadata-driven-data-quality/>
- Atlan. (2025c). *What is a data catalog? definition, features & why it matters in 2025* [Accessed October 2025]. <https://atlan.com/what-is-a-data-catalog/>
- CastorDoc. (2023). *Data catalog roi - a primer* [Accessed October 2025]. <https://www.castordoc.com/blog/data-catalog-roi-a-primer>
- ClickLearn. (2025). *40+ key statistics on digital adoption (2025)* [Accessed October 2025]. <https://www.clicklearn.com/blog/key-statistics-on-digital-adoption/>
- Databricks. (2025a). *Databricks feature store* [Accessed October 2025]. <https://docs.databricks.com/aws/en/machine-learning/feature-store/>
- Databricks. (2025b). *MLflow for ml model lifecycle* [Accessed October 2025]. <https://docs.databricks.com/aws/en/mlflow>
- Databricks. (2025c). *What is a medallion architecture?* [Accessed October 2025]. <https://www.databricks.com/glossary/medallion-architecture>
- Databricks Unity Catalog SME. (2025). *Approaches to managing unity catalog and data governance for small to mid-size teams* [Accessed October 2025]. <https://medium.com/databricks-unity-catalog-sme/approaches-to-managing-unity-catalog-and-data-governance-for-small-to-mid-size-teams-d10ca3885481>
- Dataversity. (2023). *Achieving smoother, quicker data modeling* [Accessed October 2025]. <https://www.dataversity.net/articles/achieving-smoother-quicker-data-modeling/>
- Decube. (2025). *Data catalog roi explained* [Accessed October 2025]. <https://www.decube.io/post/data-catalog-roi>
- DSSstream Team. (2025). *Unity catalog and volumes: A data engineer's perspective on modern data governance in databricks* [Accessed October 2025]. <https://www.dsstream.com/post/unity-catalog-and-volumes-a-data-engineers-perspective-on-modern-data-governance-in-databricks>
- Eckerson Group. (2024). *Data catalog adoption trends - eckerson group survey* [Accessed October 2025]. <https://www.eckerson.com/articles/preliminary-survey-results-on-data-catalog-adoption>
- EM360Tech. (2025). *7 risks of shadow it and how to mitigate them* [Accessed October 2025]. <https://em360tech.com/tech-articles/7-risks-shadow-it-and-how-mitigate-them>
- Enterprise Knowledge Team. (2025). *Three pillars of successful data catalog adoption* [Accessed October 2025]. <https://enterprise-knowledge.com/three-pillars-of-successful-data-catalog-adoption/>
- Huntress. (2025). *What is shadow data? definition, risks & prevention guide* [Accessed October 2025]. <https://www.huntress.com/cybersecurity-101/topic/what-is-shadow-data>
- Impetus Technologies. (2025). *Seamless migration to databricks unity catalog transforms data governance and accelerates ai innovation* [Accessed October 2025]. <https://www.impetus.com/resources/>

- case-studies/seamless-migration-to-databricks-unity-catalog-transforms-data-governance-and-accelerates-ai-innovation/
- Informatica Team. (2025). *4 ways change management is key to effective data governance adoption* [Accessed October 2025]. <https://www.informatica.com/blogs/4-ways-change-management-is-key-to-effective-data-governance-adoption.html>
- Integrate.io. (2025). *Salesforce data integration roi figures – 50 statistics every business leader should know in 2025* [Accessed October 2025]. <https://www.integrate.io/blog/salesforce-data-integration-roi-figures/>
- Kubrick Group. (2025). *Implementing databricks unity catalog* [Accessed October 2025]. <https://www.kubrickgroup.com/what-we-do/case-studies/databricks-unity-catalog>
- Microsoft Learn. (2025a). *Discover and manage shadow it - microsoft defender for cloud apps* [Accessed October 2025]. <https://learn.microsoft.com/en-us/defender-cloud-apps/tutorial-shadow-it>
- Microsoft Learn. (2025b). *What is the medallion lakehouse architecture? - azure databricks* [Accessed October 2025]. <https://learn.microsoft.com/en-us/azure/databricks/lakehouse/medallion>
- Onehouse. (2025). *Comprehensive data catalog comparison* [Accessed October 2025]. <https://www.onehouse.ai/blog/comprehensive-data-catalog-comparison>
- Onix. (2025). *Data governance tools compared: Unity, polaris & purview* [Accessed October 2025]. <https://www.onixnet.com/blog/data-governance-tools-a-comparative-analysis/>
- Polestar LLP. (2025). *Mastering change management for optimal data governance* [Accessed October 2025]. <https://www.polestarllp.com/blog/change-management-strategy-for-successful-data-governance>
- Rocket.Chat. (2024). *Shadow it: 6 biggest risks and how to mitigate them* [Accessed October 2025]. <https://www.rocket.chat/blog/shadow-it>
- Sandtech. (2025). *From investment to impact: A practical guide to measuring ai roi* [Accessed October 2025]. <https://www.sandtech.com/insight/a-practical-guide-to-measuring-ai-roi>
- Sentra. (2025). *What is shadow data? examples, risks & how to detect it* [Accessed October 2025]. <https://www.sentra.io/blog/securing-shadow-data>
- Sync Computing. (2025a). *The easy and comprehensive guide to understanding databricks pricing* [Accessed October 2025]. <https://synccomputing.com/the-easy-and-comprehensive-guide-to-understanding-databricks-pricing-how-it-works-and-how-to-reduce-your-cost/>
- Sync Computing. (2025b). *What is databricks unity catalog (and should i be using it)?* [Accessed October 2025]. <https://synccomputing.com/what-is-databricks-unity-catalog-and-should-i-be-using-it/>
- Turning Data Into Wisdom. (2025). *Change management for data initiatives: Ensuring success beyond technology* [Accessed October 2025]. <https://www.turningdataintowisdom.com/change-management-for-data-initiatives-ensuring-success-beyond-technology/>
- Unity Catalog. (2025). *Building an ml and ai data catalog with unity catalog and mlflow* [Accessed October 2025]. <https://www.unitycatalog.io/blogs/machine-learning-data-catalog>