



TEMPERATURE-AWARE ARCHIVAL: AGENTIC AI FOR INTELLIGENT CONTENT STORAGE TIERING ACROSS ECM PLATFORMS

Large enterprises store vast amounts of content, much of which eventually becomes inactive “dark data,” driving unnecessary storage costs and carbon emissions. This paper proposes a **Temperature-Aware Archival Framework powered by Agentic AI** that continuously analyzes usage patterns, metadata, business context, and retention requirements to predict access behavior and recommend optimal storage tiers.

Specialized AI agents automate data analysis, decision-making, policy enforcement, analytics, and governance, while maintaining human oversight for compliance-sensitive actions. The framework can reduce storage costs by **20-40%**, improve ESG outcomes, and enable scalable, compliant, and auditable content lifecycle management.

Introduction

What Is ECM?

Enterprise Content Management (ECM) is the discipline of capturing, managing, storing, and delivering unstructured content (documents, records,

correspondence, media) at an organizational scale. ECM platforms span on-premises systems (OpenText, IBM FileNet, Documentum), cloud-native object stores (AWS S3, Azure Blob, Google Cloud Storage), and hybrid environments combining both.



The Problem

Most ECM deployments treat storage as a flat cost: every document sits on the same tier regardless of value or access frequency. Organizations overpay to maintain content that may never be retrieved again. Rule-based lifecycle policies, defined once at deployment, cannot adapt to real-world usage patterns.

Why ECM Tiering Is Hard

Dark data: 80–90% of content is never accessed post-ingestion yet remains on premium storage.

Model drift: Access patterns shift with regulatory cycles and business events, degrading predictions over time.

Metadata inconsistency: Varying quality across platforms undermines tiering accuracy.

Recall latency: Archive retrieval can breach SLAs without phased rollout and fast-recall strategies.

Approval bottlenecks: Compliance-sensitive moves require human review; low-risk content needs auto-approval paths.

Platform fragmentation: Content spans systems with different APIs, schemas, and tiering capabilities.

Scale: Millions of daily objects make manual enforcement impossible.

This paper presents an agentic AI framework that predicts access likelihood and enforces tiering autonomously across heterogeneous ECM environments, cutting costs 20–40% without compromising accessibility or compliance.



AI-Powered Tiering and Retention

Static lifecycle policies cannot keep pace with the volume, variety, and velocity of enterprise content. An AI-powered approach replaces rigid schedules with continuous, data-driven decisions that align storage cost, compliance, and sustainability at scale.

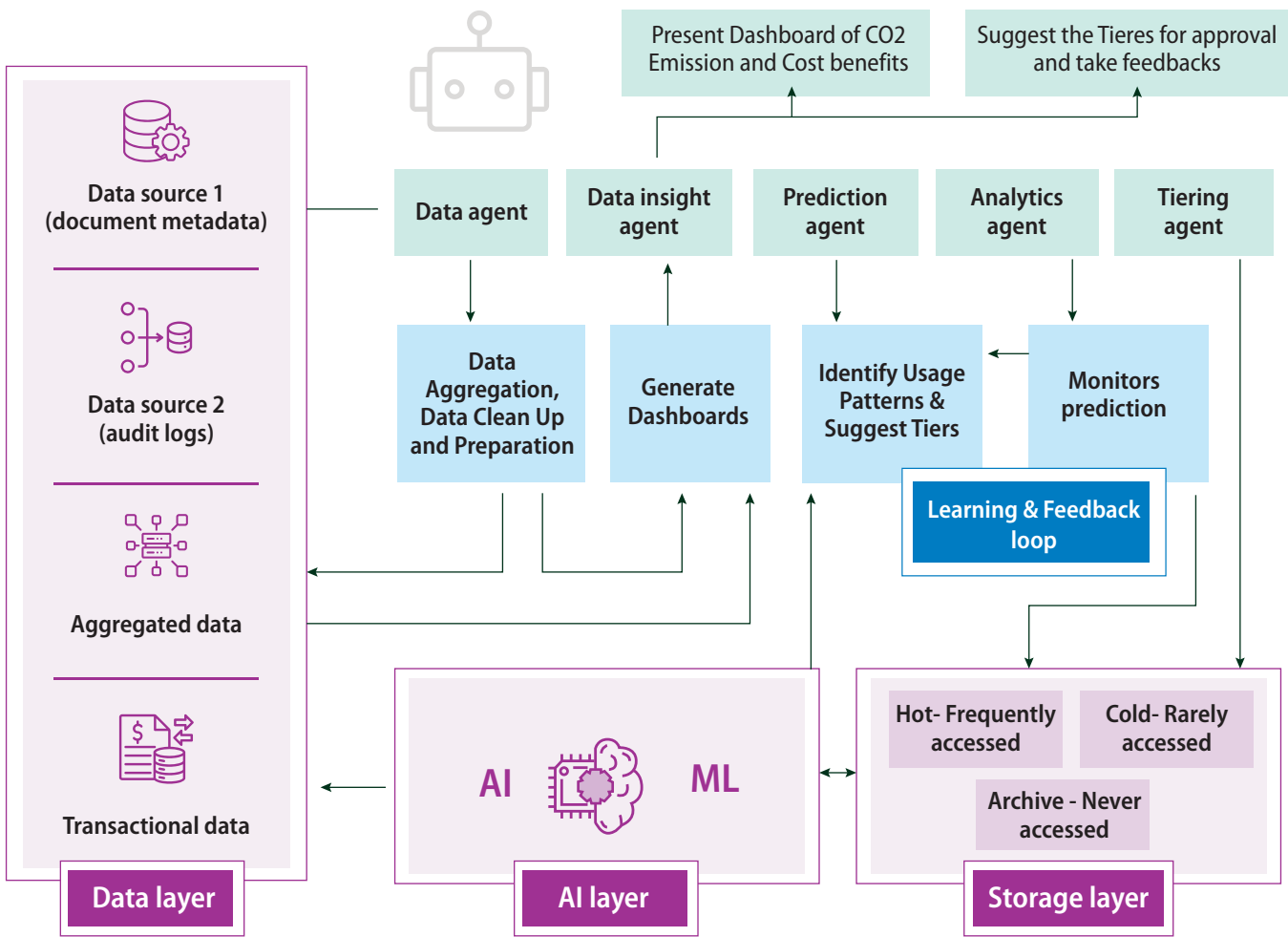
The proposed framework combines machine learning, metadata analysis, and access to telemetry to predict content placement and lifecycle actions. By learning from access patterns, content attributes, and user behavior, it dynamically transitions data across storage tier, from high-performance to deep archive, placing every object where it belongs, when it belongs there.

Proposed Technical Architecture

The proposed solution is built on a modular, agent-driven architecture designed for the scale, complexity, and governance demands of modern ECM environments, whether on-premises, cloud, and hybrid. It unifies AI-powered storage tiering with usage-aware retention management to optimize cost, sustainability, and compliance in a single framework.

Specialized agents operate across discrete layers: ingesting signals, making tiering decisions, enforcing actions, and maintaining full traceability.

Reference architecture using Azure as the cloud provider.

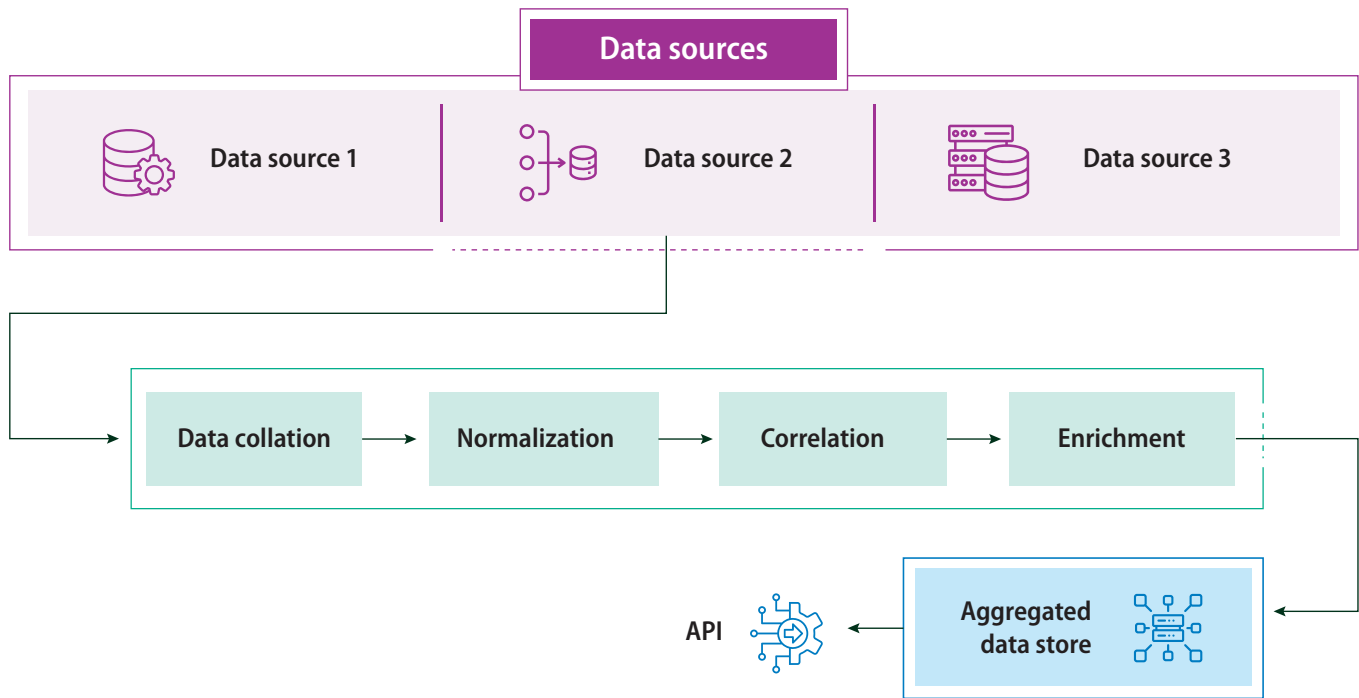


Note: Data Agent, Prediction Agent, Tiering Agent work in sequence while Analytics Agent and Data Insight agent act independently in a loop

1. Data Agent: Telemetry & Metadata Fusion

The Data Agent is the foundational layer of the proposed architecture. It ingests, normalizes, and enriches metadata and telemetry from heterogeneous

sources (content management systems, storage platforms, audit logs, and business systems) to build a unified lifecycle profile for every content object.



Core Responsibilities

 Capability	 Description
Collation	Ingests raw metadata (document ID, type, creation/modification dates) and access telemetry via APIs, change feeds, or streaming pipelines across all connected ECM and storage platforms.
Normalization	Standardizes disparate schemas into a consistent structure keyed by a unique object identifier (document ID or object URI).
Correlation	Links to access events, audit logs, and metadata records across sources to build a cohesive per-object profile.
Enrichment	Computes derived attributes: access frequency, last-access timestamp, sensitivity classification, and business ownership.
Persistence	Stores enriched object profiles in a centralized, queryable data store optimized for downstream agent consumption.
Access Interface	Exposes APIs, event streams, and batch exports for downstream agents to retrieve current enriched metadata.

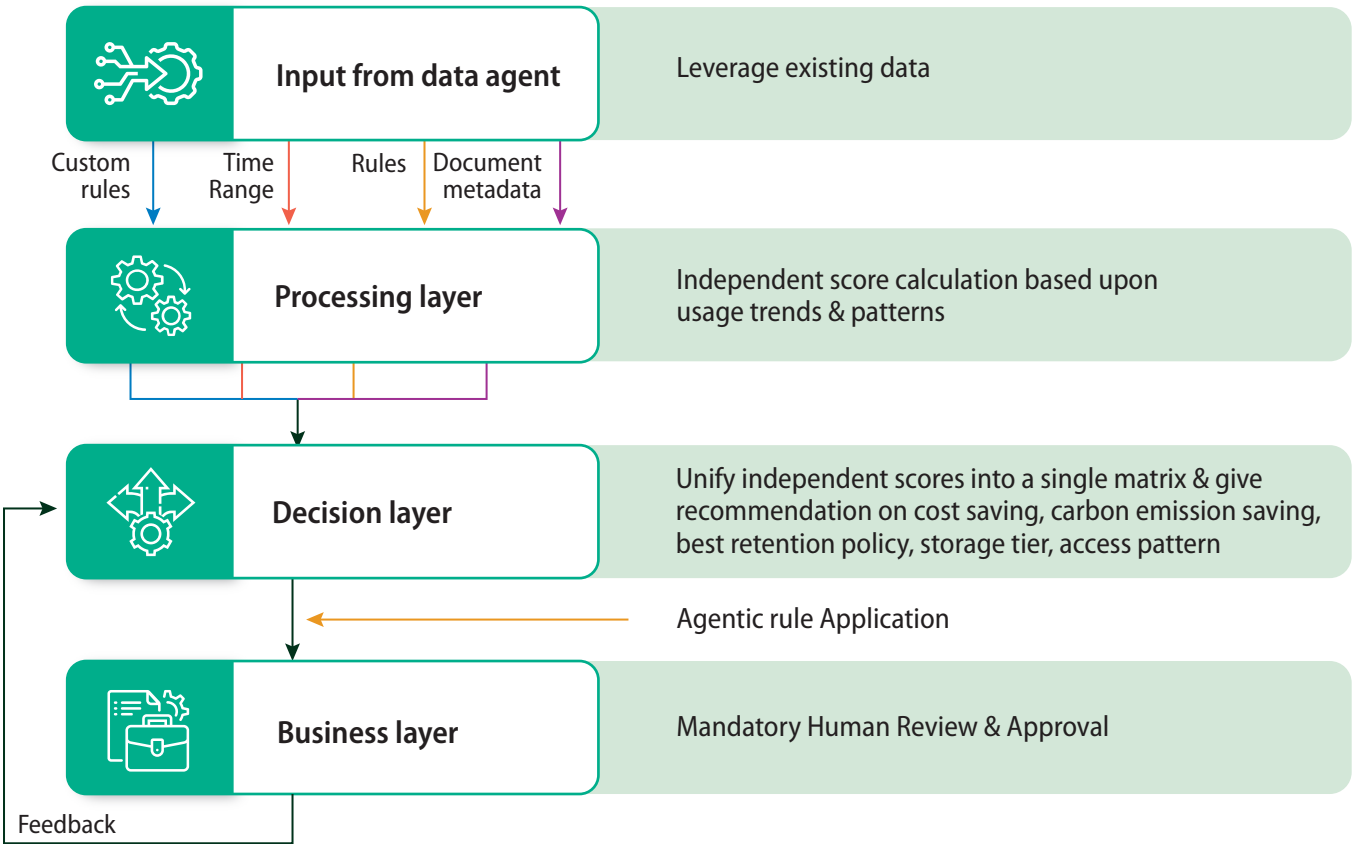
The Data Agent’s output, a clean, queryable dataset of object lifecycle profiles, serves as the single source of

truth that all downstream intelligence and policy agents depend on.

2. Insight Agent: Intelligence & Decisioning

The Insight Agent is the decision-making core of the proposed architecture. It combines machine learning, rule-based logic, and human oversight to predict access

likelihood, recommend optimal storage tiers and retention periods, and ensure every decision is explainable and auditable.



Core Responsibilities

 Capability	 Description
Classification Leverage	Uses pre-assigned document types, subtypes, and metadata tags as foundational inputs, with no reclassification required.
Access Pattern Analysis	Analyzes historical access logs segmented by document type, source, and time to identify usage trends.
Predictive Scoring	Combines metadata attributes (type, sensitivity, size), usage of telemetry, and temporal trends to score each object's future access probability.
Rule Enforcement	Applies compliance and business rules (legal holds, PII flags, retention mandates) that override or refine AI-driven recommendations.
Human-in-the-Loop	All initial tiering and retention recommendations require mandatory human approval. As confidence builds, similar objects are auto-approved via a feedback-driven learning loop.
Explainability	Maintains transparent decision records linking metadata, model inputs, rule logic, and override rationale for audit and governance.

Predictive Model Stack:

The Insight Agent selects models based on data maturity, volume, and pattern complexity:

Access Forecasting (Time-Series)

Model	Strength	When to Use
ETS (Exponential Smoothing)	Fast, low maintenance	Short history, low complexity. Default starting point
SARIMAX	Captures trends + external factors	Weekly patterns, autocorrelation detected via ACF/PACF
Prophet	Handles seasonality and holidays	Long history, strong seasonal trends
TBATS	Multiple seasonality's	High-volume objects with complex cyclical patterns

Metadata-Based Prediction (Regression)

Model	Strength	When to Use
RandomForestRegressor	Robust, interpretable	Moderate data volumes
LightGBM	High accuracy, fast	Large catalogs with sparse categories

Tier & Retention Decisioning

Predictions feed into a configurable scoring system that ranks objects by business value, access frequency, and policy constraints:

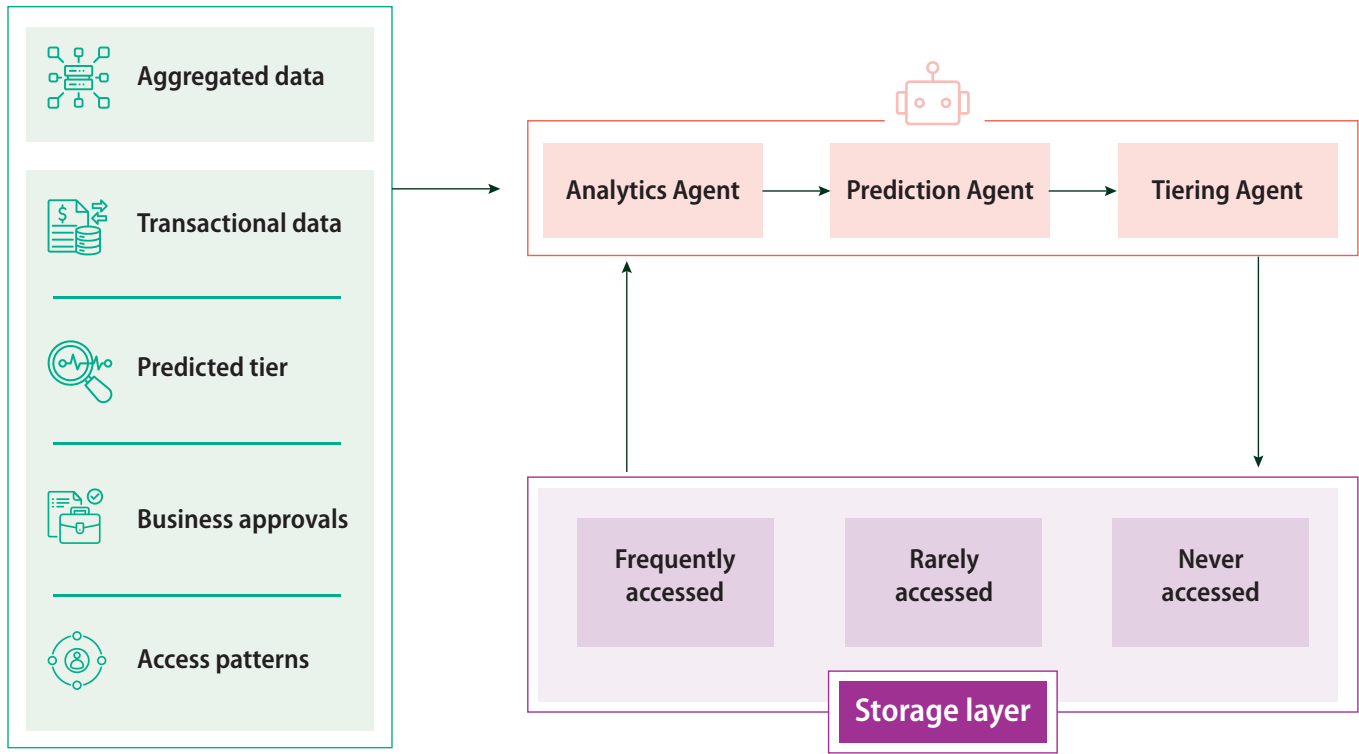
Method	Approach	Best For
Rule-Based Weighted Score	Policy-driven, transparent	Compliance-heavy environments
Cost-Based Optimization	Minimizes cost and CO ₂ impact	Environments with accurate cost models
XGBoost Ranker (LambdaMART)	Learns from labeled tier preferences	Organizations with historical tier assignment data



3. Analytics Agent: Quality Enforcement

The Analytics Agent validates the effectiveness of tiering decisions by continuously monitoring post-tiering access patterns. It acts as the feedback loop in the proposed

agentic architecture, comparing predicted usage against actual behavior to assess whether content is tied appropriately.



Core Responsibilities

 Capability	 Description
Post-Tiering Behavior Analysis	Tracks document access after tiering, identifying anomalies such as frequent retrieval of down-tiered content.
Prediction Validation	Evaluates tiering accuracy by correlating Insight Agent predictions with real-world access trends and human approval outcomes.
Retiring Recommendations	Flags content for re-evaluation or promotion to a higher-performance tier when observed usage deviates from predictions
Feedback Loop	Feeds validated outcomes back to the Insight Agent to support model retraining and reduce drift.

By closing the loop between prediction and real-world usage, the Analytics Agent enables continuous

improvement, adapting the system dynamically to evolving access patterns and governance requirements.

4. Tiering Agent: Policy Enforcement

The Tiering Agent executes lifecycle actions based on the Insight Agent’s recommendations and business policies. It enforces storage tier transitions and retention rules in a

consistent, auditable, and automated manner, handling exceptions and recalls across all connected ECM and storage platforms.

Core Responsibilities

 Capability	 Description
Lifecycle Management	Automates content transitions across storage tiers, from high-performance to deep archive, based on policy decisions. Includes down-tiering inactive content and promoting frequently accessed content as needed.
Retention Enforcement	Applies retention policies including legal holds, scheduled deletion, and compliance-driven retention periods.
Metadata & State Management	Updates storage metadata and tags to reflect current tier and retention status across platforms.
Recall & Exception Handling	Triggers immediate re-tiering or recall on demand (user retrievals, audits) and manage exceptions such as legal holds or sensitivity-blocked transitions.
Audit Logging & Notification	Logs all enforcement actions with timestamps, correlation IDs, and decision context. Notifies stakeholders for compliance and operational awareness.

The Tiering Agent translates intelligent decisions into actionable storage and retention of state changes,

optimizing costs, ensuring compliance, and maintaining operational efficiency.



5. Reporting Agent: Observability & Governance

The Reporting Agent delivers visibility, transparency, and governance across the proposed tiering and

retention of lifecycle. It aggregates operational metrics, compliance data, and decision logs into actionable insights for IT, compliance, and business stakeholders.

Core Responsibilities

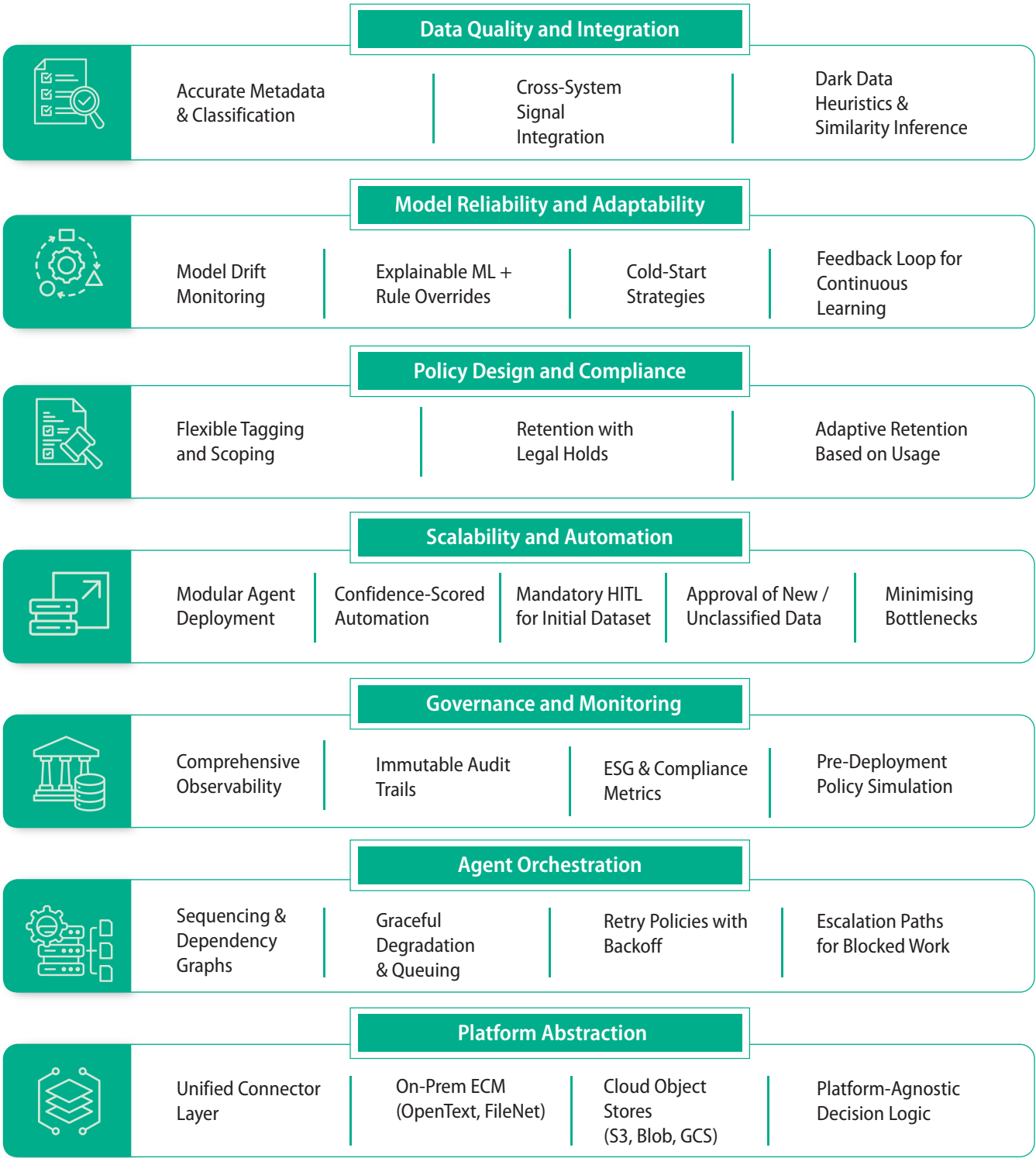
 Capability	 Description
Monitoring & Metrics	Continuously aggregates telemetry from storage systems, decision engines, and enforcement actions. Tracks KPIs including cost savings, carbon footprint (CO ₂ per TB), recall latency, SLA adherence, tier distribution, and human approval/override rates.
Auditability & Traceability	Maintains immutable logs of model versions, rule sets, policy changes, Insight Agent decisions, human overrides, and Tiering Agent enforcement actions.
Compliance Reporting	Generates reports aligned with regulatory frameworks (GDPR, HIPAA, etc.) covering retention status, legal hold compliance, PII handling, and policy version history.
Dashboards	Presents intuitive visualizations for drill-down into data trends, anomalies, and operational health, tailored to IT, compliance, and business audiences.
Governance Integration	Supports optional integration with external data catalog and governance platforms for enhanced lineage tracking and policy enforcement visibility.



Implementation Considerations

Deploying the proposed AI-driven storage tiering and retention solution in enterprise ECM environments requires careful planning across multiple domains, regardless of

whether the underlying infrastructure is cloud, on-premises, or hybrid.



1. Data Quality & Integration

Ensure accurate metadata and classification to avoid mislabeling risks. Integrate signals from distributed systems (audit logs, repositories, business applications) to build a complete data picture. For dark data or sparse logs, apply fallback heuristics or similarity-based inference.

2. Model Reliability & Adaptability

Continuously monitor for model drift and retrain as access patterns evolve. Combine ML with transparent rules for explainability and compliance. Use heuristics for cold-start scenarios and implement feedback loops from real-world outcomes to refine predictions over time.

3. Policy Design & Compliance

Use flexible tagging to scope lifecycle and retention policies. Enforce legal holds and compliance tags to prevent premature tier transitions or deletion. Adapt retention thresholds based on usage trends to balance cost against regulatory requirements.

4. Scalability & Automation

Deploy agents modularly for independent scaling and evolution. Apply Human-in-the-Loop (HITL) reviews for initial and unclassified data. Automate lifecycle actions

using confidence scoring and platform-native APIs to reduce manual effort and improve accuracy.

5. Governance & Monitoring

Enable observability via dashboards for tracking cost, performance, and compliance with KPIs. Maintain immutable audit trails for all decisions and overrides. Incorporate ESG metrics (CO₂ per TB) and simulate policy impacts pre-deployment to validate outcomes.

6. Agent Orchestration

Define sequencing, dependency, and failure-handling strategies across agents. Ensure graceful degradation: if one agent is unavailable, upstream and downstream agents queue work rather than fail silently. Establish retry policies and escalation paths for blocked transitions.

7. Platform Abstraction

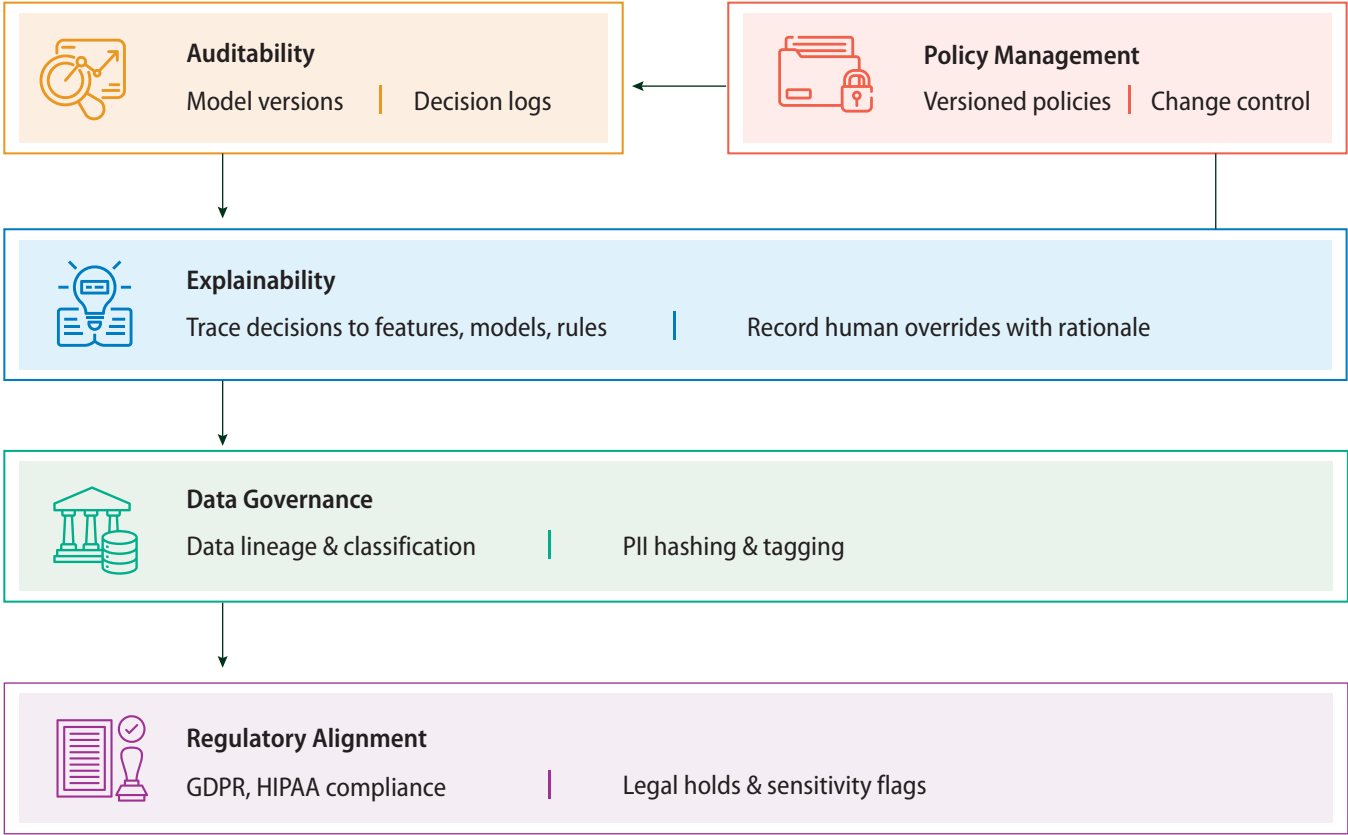
Abstract ECM and storage platform APIs behind a unified connector layer. This enables the framework to operate consistently across on-premises systems (OpenText, Documentum, FileNet) and cloud object stores (AWS S3, Azure Blob, GCS) without platform-specific logic in the decision or enforcement layers.



Compliance & Governance

Effective compliance and governance are critical to ensure the proposed AI-driven tiering and retention solution meets regulatory requirements and maintains

enterprise trust. The framework addresses this through five interconnected layers:



Benefits

The proposed AI-powered storage tiering and retention solution delivers measurable advantages across cost, sustainability, compliance, operations, and user experience.

1. Significant Cost Savings

Reduce storage expenses by 20–40% through automated, data-driven placement of content across optimal storage tiers, on-premises and cloud alike. Minimize operational overhead by automating lifecycle management and retention enforcement.

2. Enhanced Sustainability

Lower carbon footprint per terabyte by optimizing content placement on energy-efficient storage tiers. Support enterprise ESG initiatives through measurable reductions in storage-related emissions.

3. Improved Compliance and Risk Management

Enforce retention policies dynamically, ensuring content is retained only as long as required by regulations and internal mandates. Mitigate risks of accidental deletion or improper tiering with human-in-the-loop approvals and audit-ready logs.

4. Operational Efficiency and Scalability

Automate complex decisions that are impossible to manage manually at scale, especially with millions of daily content ingestions. Enable a modular, scalable architecture that adapts to evolving business needs and growing data volumes.

5. Better User Experience and SLA Adherence

Maintain retrieval of SLAs by intelligently predicting access patterns and preemptively positioning content. Reduce latency and avoid costly recall penalties with proactive lifecycle management.

Limitations

While the proposed agent-driven architecture provides scalable and intelligent storage tiering, several limitations should be acknowledged:

- **Cold Start & Sparse Data Gaps:** For newly ingested or rarely accessed content, the system may lack sufficient telemetry for confident tiering decisions, falling back to heuristics that may not capture edge cases.
- **Metadata Dependency:** Prediction accuracy and policy enforcement depend on high-quality metadata. Incomplete or inconsistent tagging can lead to misclassification, compliance risks, or unnecessary storage costs.
- **Model Drift & False Positives:** Despite feedback loops, access patterns can shift unpredictably, leading to incorrect tiering. Retraining lag or overfitting may exacerbate this.
- **Human Review Bottlenecks:** HITL review for unclassified or high-risk content can introduce delays, particularly during large-scale onboarding or compliance audits.
- **Integration Complexity:** Correlating data from multiple ECM platforms, log systems, and access monitors may require custom connectors that increase maintenance overhead.
- **Regulatory Sensitivity:** Automated retention and deletion actions must be carefully governed to avoid non-compliance with evolving privacy regulations (GDPR, HIPAA). Misconfigurations can have legal consequences.
- **Environmental Cost of Automation:** Continuous analytics, model retraining, and large-scale processing introduce their own carbon footprint if computing resources are not optimized.

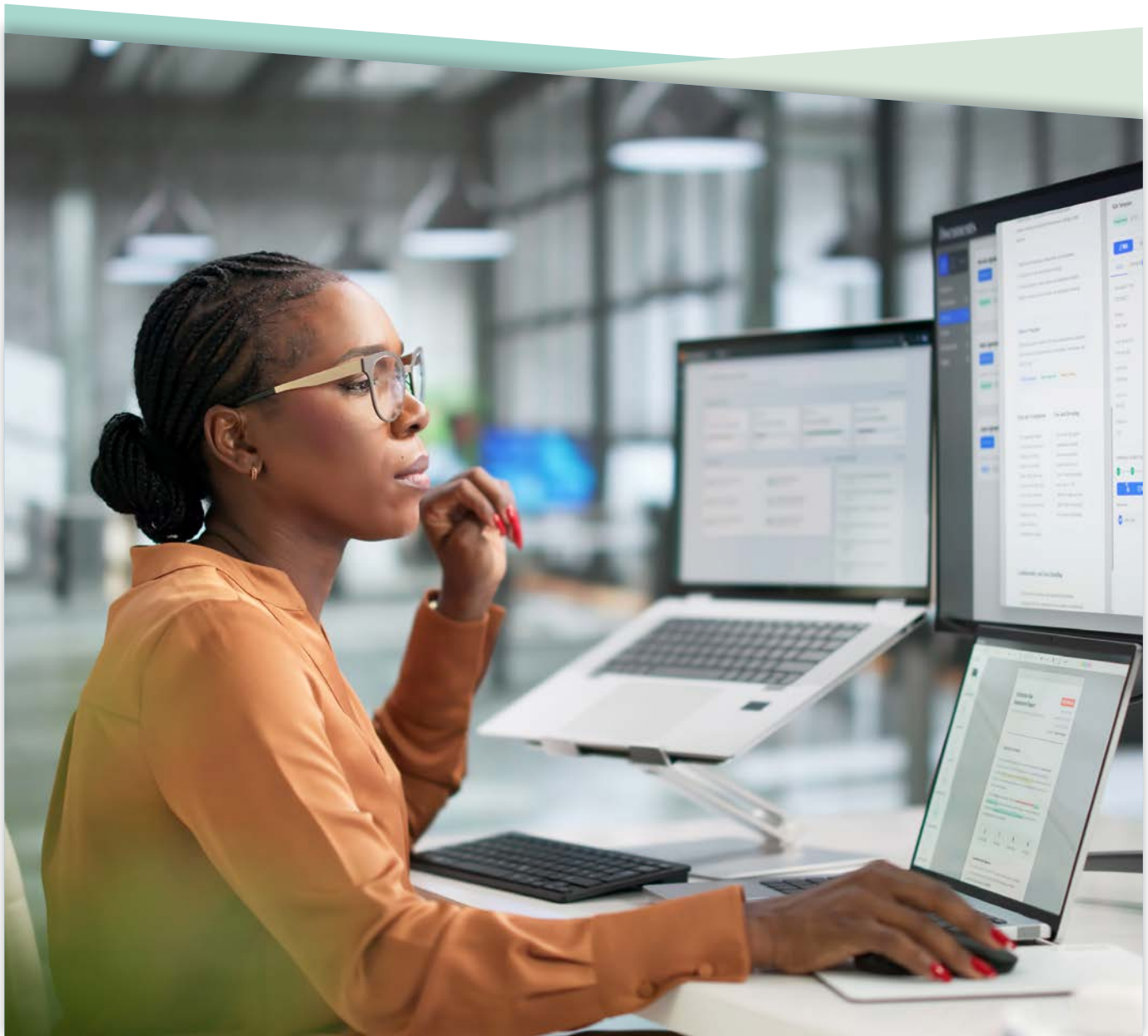
Conclusion

Enterprise content is growing faster than any manual process can govern, and every terabyte stored on the wrong tier translates into unnecessary cost, carbon emissions, and risk from data that may never be accessed again. Static lifecycle rules were designed for a different era; they cannot assess content value, predict future access patterns, or confidently justify archival and deletion decisions during audits.

The **Temperature-Aware Archival Framework** transforms these economics. By combining predictive intelligence,

autonomous agents, and human oversight, it ensures that every piece of content resides in the most appropriate storage tier at the right time, turning storage from a growing liability into an audit-ready and sustainable asset. Machine intelligence handles millions of routine decisions, while human judgment governs the critical exceptions.

In an era of exponential data growth and increasing regulatory scrutiny, the question is no longer whether storage should become intelligent, but how quickly organizations can make the transition.



About the Authors



Ishwarya Kumar,

Technology Architect with Digital Practice Unit in Infosys

Bio: Ishwarya Kumar is a Technology Architect with 15+ years of experience in enterprise application design and delivery, specializing in Enterprise Content Management, Intelligent Document Processing, and Generative & Agentic AI on Microsoft Azure. She architects AI-driven content solutions that bring automation, intelligence, and scalability to large-scale document ecosystems.



Aishwarya Sundaresan,

Lead Consultant with Digital Practice Unit in Infosys

Bio: Aishwarya Sundaresan is a Lead Consultant at Infosys with over 15 years of experience delivering enterprise technology and digital transformation initiatives. She currently specializes in Agentic AI, AI-powered SDLC, Generative AI, LLM applications, and RAG-based solutions. Her work focuses on helping organizations adopt AI responsibly while balancing scalability, cost, performance, and business values. With a foundation in full-stack development using Java, Python, AWS, and Agile DevOps practices, she brings a practical, business-focused approach to solving complex enterprise challenges.



Aswini Chandrasekaran,

Technology Architect with Digital Practice Unit in Infosys

Bio: Aswini Chandrasekaran is a Technology Architect with deep expertise in OpenText enterprise content platforms and Generative & Agentic AI. She designs intelligent content solutions that integrate AI-driven automation with enterprise-grade ECM, helping organizations modernize and unlock the full value of their content.

For more information, contact askus@infosys.com



© 2026 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.