CAEVES

Reinventing Legal Discovery with AI-powered Deep Storage

Executive Summary

Legal and compliance teams have faced the same enduring challenge for decades: how to surface relevant information from decades of archived, unstructured data when discovery or regulatory review demands it. Traditional archives retain data but fail to make it accessible or meaningful. The result? Delayed responses, mounting costs, compliance risks, and inefficient reliance on outside vendors. The legal discovery process, once a safeguard for governance, has too often become a liability. Businesses find themselves floundering. New requirements, like the EU Data Act, further muddy the waters.

This paper examines the challenges these businesses face, their origin, and the best way to resolve them. Information present in this paper is based on the experience of CAEVES™ engineers and leadership, as well as direct testing in customer environments.

We've seen that CAEVES AI-powered Deep Storage resolves many of these challenges. By combining cloud-native data lakes in Microsoft Azure with Agentic AI search and retrieval, CAEVES transforms archived data from static cost centers into intelligent, searchable assets. The benefits are immediate: faster case responses, defensible accuracy, reduced spend, and measurable ROI across legal and compliance operations.

> "Global AI regulations vary wildly...complicating alignment of AI investment with demonstrable and repeatable enterprise value and possibly opening enterprises up to other liabilities."
>
> Lydia Clougherty Jones, Sr. Director Analyst at Gartner

Modern Challenges of Legal Discovery

Despite decades of investment in digital archiving and legal discovery tools, processes, and workflows, most enterprises still struggle with the same four underlying challenges in legal discovery:

ᐱ Inaccessible Data: Slows discovery timelines and forces legal staff to rely on overworked IT teams or external vendors to extract and process information

ᐱ Reactive Processes: Leaves little time for context or accuracy, making teams scramble to find data scattered across email archives, backup systems, and cloud platforms.

ᐱ Excessive Vendor Costs: Data frequently leaves internal environments for third-party processing, where each transfer introduces new costs, security considerations, and workflow bottlenecks.

ᐱ Limited Accuracy: Accuracy gaps force legal professionals to over-collect and over-review, driving up both cost and cycle time.

Despite decades of investment in eDiscovery platforms, content management systems, and compliance frameworks, systemic issues remain entrenched. Most organizations have modernized around the edges—adding new analytics tools, workflow automation, and dashboard visibility—without addressing the underlying problem: the archive itself.

Most legacy archives were built for long-term retention, not accessibility. Over time, files become trapped in proprietary formats or offline storage, making even basic search and retrieval a complex, manual exercise. Data volumes continue to grow exponentially, drawing even more regulatory scrutiny and concerns about data breaches and privacy.

When legal teams attempt to apply modern discovery tools to this legacy foundation, they encounter an immediate bottleneck. Data locked away in proprietary formats, offline storage, or fragmented repositories can't be indexed effectively or queried at scale. Each retrieval becomes an extraction project; slow, manual, and often dependent on IT intervention or third-party vendors.

Over time, enterprises respond by layering new technologies onto the same brittle architecture. Search engines, compliance modules, and review platforms are added in the hopes of improving speed and visibility. But without seamless access to the underlying data, these tools can only provide incremental gains. The result is an ecosystem that looks modern on the surface but remains opaque, siloed, and fragile at its core.

Sovereignty and Strategy

Legal disputes related to the introduction of AI in legal discovery are on the rise, so it's clear that the indiscriminate addition of AI into these patchwork environments will only worsen the situation. While AI does provide an incredible use case for searching and indexing files in legacy archives, improper use could open businesses up to significant lawsuits.

In a recent survey, Gartner asked 360 IT leaders about their rollout of Generative AI (GenAI). Over 70% of respondents marked regulatory compliance as a significant challenge in the process, highlighting the issues widespread GenAI usage might create for legal discovery. Whether it's for productivity assistants or AI-assisted search, it's clear that data sovereignty must be considered when building an AI solution for legal discovery.

"Global AI regulations vary widely, reflecting each country's assessment of its appropriate alignment of AI leadership, innovation and agility with risk mitigation priorities," said Lydia Clougherty Jones, Sr. Director Analyst at Gartner. "This leads to inconsistent and often incoherent compliance obligations, complicating alignment of AI investment with demonstrable and repeatable enterprise value and possibly opening enterprises up to other liabilities."

In the past year, two major firms, Midjourney and D5 Render, have made attempts to adjust their EULAs to automatically claim the rights to customer-created data. Other companies, like Autodesk, have prohibited the use of any models created in their software to train AI, even if the process is entirely in-house. Even more vendors have been adjusting their EULAs to account for the growth of AI, so it's clear that businesses are taking a stand and that AI strategy needs to involve more than the blanket introduction of GenAI into business environments.

Well-designed data compliance and governance strategies must form the foundation for implementing AI within a system designed to protect customer data and accommodate these evolving EULAs. The most obvious point of reformation is the data itself.

Compliance with Deep Storage

Compliance and discovery are mostly reactionary functions triggered by audits, investigations, or legal events. Finding a new process that still maintains compliance requires shifting perspective, processes, and platforms.

The EU Data Act, enacted to create a fair and innovation-ready data economy, underscores this transformation. It introduces a new paradigm where data accessibility, portability, and interoperability are not optional. Organizations are now expected to make data available to those entitled to use it, ensure transparency in how data is managed, and maintain the ability to trace its use across the lifecycle. This builds on previous EU regulations, such as the GDPR, further protecting consumer data.

This regulatory shift means compliance is no longer measured solely by the ability to retain or produce data, but by how effectively an organization can access, interpret, and share it under governed conditions. In other words, compliance is moving from passive retention to active readiness, and traditional archive infrastructures weren't designed for that.

Deep storage offers a new dynamic by simplifying accessibility on a new platform. With it, active and archived data are unified in the cloud and customer data is not exposed to the risks of on-premise storage, fractured data silos, or unmonitored legacy archives. Deep storage makes accessibility the default, not the exception.

It unifies active and archived data under a search-first architecture where every object—structured or unstructured—is indexed, discoverable, and context-aware. Instead of extracting or restoring data to external systems, legal and compliance teams can locate and review information directly within a centralized, queryable environment.

This model reduces the inherent risks of unmonitored or ungoverned storage—such as accidental exposure, data loss, or incomplete audit trails—while ensuring that sensitive information remains fully controlled within the organization's governance framework.

By design, deep storage:

⌃ Simplifies Accessibility: Active and archived data coexist in one environment, eliminating the friction between storage tiers and reducing dependence on manual retrieval.

⌃ Enhances Governance: Built-in indexing and metadata mapping create a transparent record of what data exists, where it resides, and how it has been accessed or modified.

⌃ Improves Security Posture: Cloud-based architecture mitigates the risks associated with fragmented on-premise archives, ensuring consistent policy enforcement and encryption across the data lifecycle.

The Strategic Imperative

Three converging forces make the transition to deep storage urgent and unavoidable:

⌃ Exploding Data Growth: Enterprise data continues to expand exponentially across sources and formats, making legacy archives increasingly unmanageable and opaque.

⌃ AI Maturity: Modern AI systems, especially RAG and LLM-driven applications, depend on historical, contextual data to operate effectively. Without accessible archives, AI operates with blind spots.

⌃ Regulatory Pressure: Global mandates such as the EU Data Act are redefining expectations around transparency, portability, and stewardship. Compliance is no longer just about retention; it's about continuous control and verifiable readiness.

Together, these factors demand a new model for enterprise data governance; one that unifies accessibility intelligence, and compliance under a single architecture.

CAEVES: Locked Archives to Living Records

For decades, enterprise archives have functioned as digital vaults—built for retention, not readiness. They've met compliance obligations but done little to enable intelligence, discovery, or innovation. As organizations enter an era defined by AI, automation, and continuous regulation, this static approach no longer works.

As enterprises modernize their data strategies, the gap between traditional archiving and intelligent deep storage becomes increasingly clear. By simplifying how information is indexed, searched, and retrieved, deep storage strengthens compliance postures across global regulatory frameworks without adding complexity or cost. CAEVES was built natively in Microsoft Azure to redefine the archive itself. Rather than relegating decades of enterprise data to inaccessible repositories, CAEVES transforms it into an AI-indexed, cloud-scale deep storage platform. Here are a few examples of CAEVES simplifies legal workflows.

The following table outline the core differentiators that define this shift and highlight how CAEVES™ AI-powered deep storage fundamentally outperforms legacy models in speed, accuracy, and strategic value.

"Compliance is no longer measured by data retention—but by accessibility and control."

Jaap van Duijvenbode, Head of Product & Customer Experience

Dimension	Legacy Archives	Deep Storage (CAEVES™)
Purpose	Retain data only	Transform and search data for intelligence and accessibility
Searchability	Minimal; metadata-only	Semantic and contextual search across structured and unstructured data
Accuracy	Manual, incomplete, and error-prone	AI-driven, explainable, and auditable
Cost Model	Vendor-heavy, pay-per-search	Internalized, ROI-driven, scalable efficiency
Speed to Results	Weeks to months	Minutes to hours, even at enterprise scale

Time to Discovery: From Weeks to Hours

Traditional discovery processes rely on manual restores, multi-vendor coordination, and extensive review cycles. CAEVES™ AI-powered Deep Storage eliminates these dependencies by making all data—active or archived—instantly searchable in a centralized environment. What once required weeks of cross-departmental effort and external support can now be accomplished in hours, allowing teams to respond to audits, investigations, or regulatory requests almost in real time.

Legal Spend: Millions in Annual Savings

Legacy discovery workflows consume vast resources in data preparation, vendor management, and outside counsel review. By enabling direct access to clean, indexed data, CAEVES™ AI-powered Deep Storage significantly reduces review volumes and minimizes redundant data handling. For large enterprises, this translates to millions in annual savings, freeing legal budgets for more strategic initiatives instead of reactive cost recovery.

Risk Mitigation: Comprehensive, Defensible Search

Incomplete or inconsistent search results are among the greatest legal risks in high-stakes litigation and regulatory review. CAEVES™ ensures that every record, structured, unstructured, or legacy, is included in the discovery scope, with rich metadata and audit trails to validate completeness. This approach strengthens defense by ensuring that responses are consistent, verifiable, and fully aligned with internal governance policies and external regulatory expectations.

Productivity and Oversight: From Reactive to Proactive

By automating the most time-consuming components of discovery and compliance, CAEVES™ allows legal and compliance teams to shift from firefighting to foresight. Staff can focus on higher-value analysis, policy refinement, and continuous monitoring rather than repetitive data retrieval. The result is a more agile compliance function, namely one that prevents issues before they escalate and maintains continuous readiness for regulatory engagement.

Conclusion

For too long, enterprises have treated archives as static vaults; ultimately necessary for compliance but inaccessible for insight. This outdated mindset has cost organizations not only time and money but also trust, agility, and strategic advantage. The hidden inefficiencies of legacy archives have quietly eroded operational confidence, leaving legal and compliance teams to navigate complex data landscapes with outdated tools and incomplete information.

Compliance, discovery, and governance no longer exist in isolation, they are now central to digital transformation. Regulations such as the EU Data Act are reshaping expectations for data accessibility, transparency, and portability, demanding that enterprises treat information as a shared, accountable asset rather than a locked repository. Meeting these new standards requires more than incremental upgrades; it requires a rethinking of how organizations manage and activate their historical data.

CAEVES™ Capability	Description
Agentic AI Search	Semantic indexing and retrieval-augmented generation (RAG) enable natural language queries and intelligent discovery.
Defensible Accuracy	Search results are comprehensive, explainable, and auditable to support regulatory and legal defensibility.
Seamless Cloud Integration	Natively integrated with Microsoft Azure governance, security, and compliance frameworks.
Proactive Compliance	Continuous tagging, classification, and monitoring to anticipate regulatory requirements and policy shifts.

CAEVES™ AI-powered Deep Storage represents that shift. By embedding Agentic AI into an Azure-native deep storage architecture, enterprises can unify decades of structured and unstructured data into a single, intelligent environment that is secure, compliant, and ready for use. This isn't simply modernization, it's transformation.

With CAEVES, organizations can:

⌃ Make decades of unstructured data instantly searchable, eliminating manual retrieval and enabling AI-driven discovery.

⌃ Respond to regulators and courts with confidence, supported by verifiable, defensible search results that meet evolving compliance standards.

⌃ Dramatically reduce legal and operational spend by cutting weeks of manual work down to hours through automation and AI-assisted review.

⌃ Elevate compliance from reactive obligation to strategic advantage, transforming oversight into a proactive, value-creating function that enhances enterprise resilience.

Enterprises that continue to rely on legacy archives risk falling behind, not just technologically, but operationally and legally. The future of compliance and discovery demands a new foundation: one built on intelligence, accessibility, and trust, where information becomes an active, intelligent layer of the enterprise.

To learn more about CAEVES, visit our website.