2025

The Rise of In-Place Data Management

ZL In-Place Data Management White Paper

ZL TECHNOLOGIES | WWW.ZLTI.COM

ZLTECH

Executive Summary

The most pressing technological initiative for large organizations is to utilize corporate data for a competitive advantage and harness it for AI. However, as the volume of unstructured data explodes, organizations face increasing challenges in managing their information and extracting value from it. Traditional archiving and analytics models that are built on duplicating and storing copies of data cannot scale to unstructured data volumes, given the resulting storage costs and governance risks.

In-Place Data Management represents a fundamental shift in how enterprises manage information. By extracting the "essence" of every document without copying the document itself, the ZL Platform enables organizations to search, "wrangle" and unlock the true value of unstructured data while managing it at the source for information governance requirements such as privacy, compliance, eDiscovery, and records management.

The Challenge of Managing Unstructured Data at Scale

Previous paradigms for managing unstructured data have become ineffective in the wake of the 21st century data explosion, growing into the tens and hundreds of petabytes at most large companies. With over 80% of the accumulating data existing in unstructured formats like messages and files, organizations face disproportionate legal, regulatory, and compliance risk while the real value remains out of reach. Created by humans, for humans, the intelligence contained in unstructured data makes it as sensitive—and difficult to manage—as it is valuable. This data contains the intent, sentiment, and dynamics of the workforce—making it a highly strategic asset for training AI.

The Brick Wall of Traditional Archiving

Many solutions have emerged over the years for various legal, regulatory, and business applications for unstructured data. Archives provide a solution for storing and retaining growing enterprise data, while Enterprise Content Management (ECM) systems exist for managing business records, and data lakes are used to store data for processing. Each of these solutions inherently duplicate data, resulting in additional storage costs and exposure to legal, regulatory, and cybersecurity risk. As companies add new repositories for these respective purposes, they create more silos that must be managed, each of which has limited communication with one another and varying capabilities for managing data. This effectively makes it impossible to execute complex functions such as managing personal data for privacy requirements, which must closely align with other governance policies. Moreover, at today's volumes of unstructured data, the cost associated with copying (potentially multiple times) all this data is unfeasible.

The result of such application limitations is that a relatively low percentage of all enterprise data is managed, while the vast majority remains beyond control. As organizations deliver on their AI initiatives, there is a need to manage *everything else,* extracting intelligence from all the unstructured data across the organization, without copying or duplicating data.

AI: Trapped in the Sandbox

Unable to access the full breadth of their unstructured data, companies have started training AI with small, incomplete data sets known as sandboxes rather than the complete, rich landscape of enterprise data, or "the beach". These sandboxes reduce reliability, as AI systems cannot acknowledge what they don't know and hallucinate to fill in the gaps when given irrelevant information or an incomplete data set. To unlock the real value of artificial intelligence, these models need access to the most relevant data combed from the entire beach, not just the sandbox.

The Fear of Deletion

Companies have long erred on the side of data retention, driven by the fear of deleting something important—whether it's a regulatory record, potential evidence in litigation, or information that could prove valuable in the future. However, the risk of over-retention has grown with the rise of Redundant, Obsolete, and Trivial (ROT) data, as have the costs associated with storing it all—particularly as data is increasingly shifted to the cloud. Over-retaining also significantly increases the risk of cybersecurity breaches, as well as legal and regulatory exposure.

Companies today are in search of a solution that would allow the "defensible" deletion of data over time, driven by policies aligned with their regulatory, legal, and business requirements.

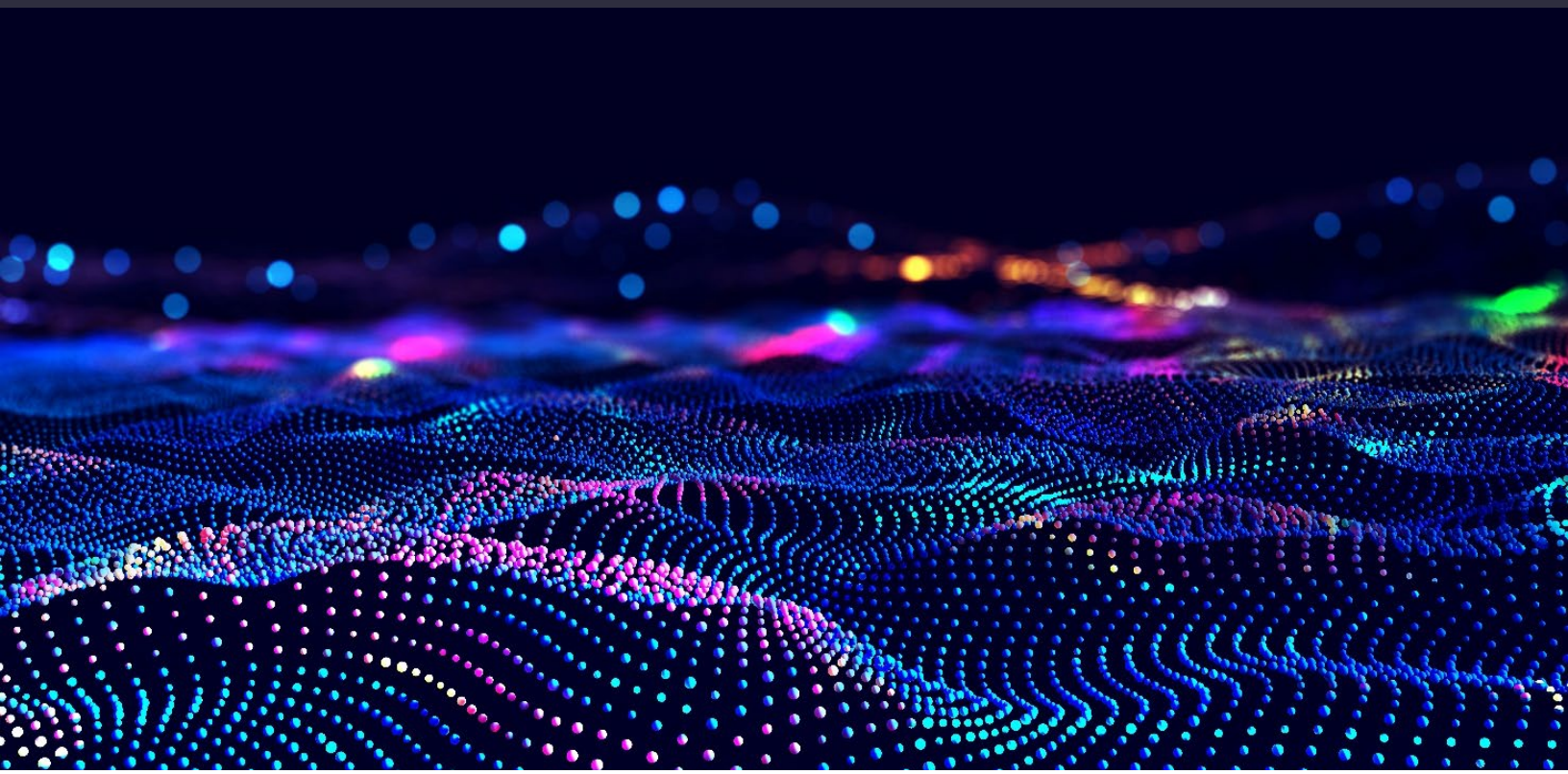
The Speed-to-Data Problem

Not only do companies need to prune data over time, they also need the capability to find and wrangle relevant information at their fingertips, requiring search architecture powerful enough to scrape every word of all unstructured data sources, spanning email, file shares, SharePoint sites, collaboration platforms, etc.

Some enterprises have come to rely on the native capabilities in their email and collaboration platforms, such as Microsoft 365. However, these systems were designed for user productivity and departmental collaboration. The resulting architecture relies on thousands of individual indexes—rather than a single master index—meaning that searches must parse through thousands of individual indexes, which significantly impairs speed and reliability, often failing to meet stringent eDiscovery and compliance requirements. Additionally, solutions that don't maintain a full-text index must reprocess the entirety of their data when regulations or corporate records definitions change.

When organizations request crucial data stored in Microsoft 365, the information extraction is "throttled," meaning there is a limit to how much can be extracted at once, which can cause exports to take hours, days, or even weeks. Encountering a standstill every time a new extraction is initiated makes timely access to data impossible.

Data volumes and governance requirements have outpaced the capabilities of native platforms. With the advent of GenAI, it's critical that organizations can **access and manage data across the entire enterprise** in an iterative and evolving capacity.



Information Governance Meets AI and Analytics

Enterprises face a major challenge when it comes to implementing AI at scale. Open-source AI platforms trained on public data cannot produce outputs that are relevant or useful for enterprise operations, nor can these platforms be trained using company data due to security concerns. The result is that organizations must develop their own AI models using their own data. Internal AI is heavily dependent on unstructured content, but it requires targeted, relevant data, which must be filtered from undesirable data such as ROT and sensitive data. Today, organizations struggle to find, cull, and deliver the most relevant data for AI from across the enterprise while ensuring data is properly managed for governance, risk, and compliance.

Poorly governed data leads to unreliable analytics and increased risk exposure with AI. Currently, there are few built-in guardrails to prevent AI from ingesting sensitive information or filtering out trivial or unreliable content. Once AI is trained, you cannot untrain it. Attempting to do so would be like trying to remove a drop of ink from a swimming pool: it cannot be extracted. What's missing today are the "governance guardrails" for unstructured data, including eDiscovery, privacy, records management and regulatory compliance.

To break through barriers encountered in traditional archiving, organizations need a fundamentally different approach to data management.

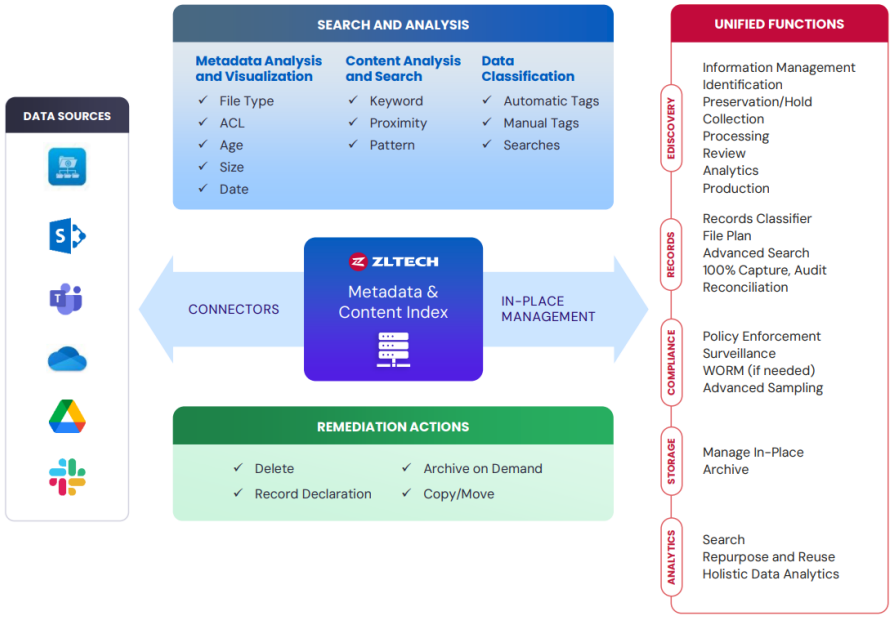
In-Place Management with the ZL Platform

In-Place Management: The Future for Unstructured Data

While organizations have historically managed approximately 5% of unstructured data for various business requirements—creating repositories for records management, archiving, and eDiscovery—today there is a need to manage the entirety of the unstructured data environment. To address the remaining 95%, the ZL Platform has been built to manage data "virtually," or in-place, in which messages and files are managed without copying the original document. The platform extracts and indexes the essence of every document, including metadata and content, while only high-value documents such as contracts are archived. This enables the capability to search, wrangle, and feed relevant, governed data to AI repositories.

From a single, unified platform, users can execute records management functions on the data source, including document classification, retention, review, and defensible deletion. Manual and automatic data tagging allow users to classify documents and apply custom policies that align with regulatory requirements. In addition, the user can declare high-value files as records, remediate personally identifiable information (PII), and flag ROT files for deletion.

In-Place Data Management allows for rapid reclassification and policy updates all from a single platform, eliminating silos and reducing inconsistency. Organizations can cut storage costs, reduce legal risk, and simplify operations by removing the logistical burden of archiving.



In-Place Data Management enables complete information governance from a single unified platform

The Whole Beach: Training AI and Conducting eDiscovery

As a result of In-Place Data Management, companies can access the full beach of enterprise data rather than incomplete sandboxes, enabling more thorough AI training and search for eDiscovery.

For example, a large financial institution sought to leverage AI to analyze and review large volumes of emails and messages for regulatory violations and other business risks. However, they were constrained by the limitations of current AI technology, which could only process up to 20,000 messages per day. Their sole method to collect 20,000 messages was to sample one sandbox at a time and hope the sample included relevant information.

These limitations created two fundamental challenges. First, they needed to search, cull, and deliver only the most relevant data for analysis, a particularly difficult task given the disorganized nature of unstructured data sprawled across the enterprise. Second, they needed to ensure that all documents selected for analysis were governed, lifecycle-managed, and filtered for sensitive content.

The financial institution turned to ZL Tech, implementing the ZL Platform to unify data management and enterprise-wide search. With ZL, they were able to efficiently comb the entire beach, allowing the institution to locate and extract relevant content in alignment with their AI processing limits and ensure that all exported data met governance, privacy, and compliance standards. This success story highlights ZL's ability to help large enterprises implement AI programs responsibly and efficiently.

ZL's In-Place Data Management can also be leveraged to streamline the eDiscovery process and gain the upper hand in litigation. Searching the entire "beach" of unstructured data with a fine-tooth comb can find an order of magnitude more "smoking guns"—pieces of evidence that can significantly impact a case in their favor—as traditional sandbox-based searches, essentially weaponizing eDiscovery. This can give a company high ground in keyword negotiations, helping to win or settle lawsuits early and advantageously. Whether organizations are conducting internal investigations, audits, or training AI, they can search across all unstructured enterprise data to find exactly what they need quickly, securely, and with confidence.

Curation Through Defensible Deletion

As companies work towards accessing the entire beach of enterprise data, they encounter a significant amount of ROT. Before it can be leveraged for insights, this data must be cleaned up and deleted defensibly—ensuring policies are aligned with each governance function and maintaining audit trails to show why data was deleted. The ZL Platform enables companies to curate, classify, and dispose of data over time, dramatically increasing the relevance of data while reducing storage costs by as much as 60%.

In practice, defensible deletion delivers substantial cleanup breakthroughs to enterprise data environments. For instance, a U.S.-headquartered global bank managing billions of documents across its global file share repositories faced escalating risk and costs from data sprawl and ROT.

Initial internal efforts to clean up data resulted in minimal progress, with just 1 million files deleted over a year. The bank lacked centralized control, which led to inconsistent retention policies across regions, and a bloated volume of ungoverned files.

In 2023, the bank adopted ZL Tech's In-Place Data Management and scaled its cleanup to 100 million documents per year. This shift to in-place governance delivered major cost savings, reduced legal and privacy risk, and laid the groundwork for broader data transformation efforts, including AI enablement and selective archiving.

Speed-to-Data

By extracting intelligence from large data repositories such as Microsoft 365, organizations avoid the "throttling" incurred when exporting large data volumes into AI and analytics repositories, accelerating time-to-data by one-thousand times or more. Microsoft is typically limited in terms of the capability it offers users to review a dataset before exporting, and the export rate itself is throttled to a rate that is often impractical for bulk exports. If further iterations are needed, the cycle must be repeated.

As data is created, the ZL Platform extracts the essence of every document, which can be searched, culled, and channeled into AI repositories in near-real time. When relevant information is identified, enterprises can easily retrieve the original document if needed to preserve it or leverage its intelligence in analytics and AI repositories.

Information Governance Unlocks AI and Analytics

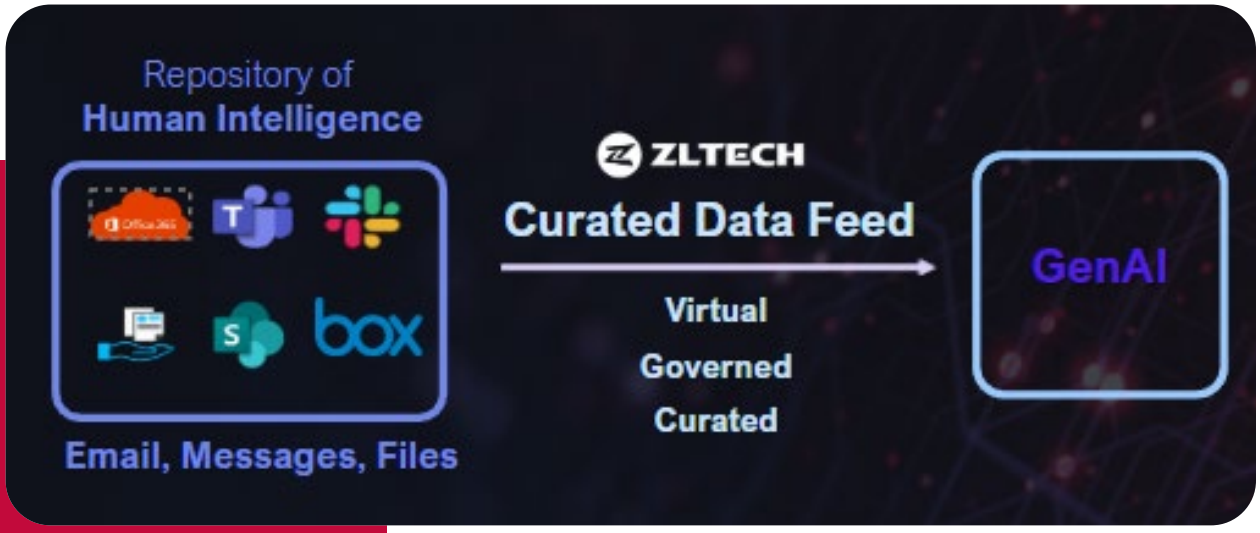
Beyond accessing vast amounts of data with speed and accuracy, ZL Platform's capabilities can be leveraged to turn the complete beach of information into valuable company insights. When surfaced, information created and shared by employees has the power to reveal a side of the enterprise never yet seen. Unstructured data reveals both the content and intent behind all employee communications. Who are the "go-to" employees? Which employees are leading initiatives behind the scenes?

In-Place Data Management allows users to curate a data set for internal AI by consolidating the enterprise data ecosystem into a "virtual" data lake. From there, they can search, cull, and deliver the most relevant information to the appropriate AI systems. Feeding relevant and complete data to AI minimizes hallucinations and improves accuracy. Department-specific applications can be trained with data tailored to their unique needs, enabling more precise results and better business outcomes.

By keeping sensitive information out of AI training sets, ZL helps reduce exposure risk for compliance, privacy, and legal obligations so organizations can harness their unstructured data with confidence.

Conclusion

ZL Tech's In-Place Data Management delivers a transformative approach to information governance, one that meets the demands of today's enterprise data volumes without the limitations of traditional archives. From defensible deletion and full-enterprise search to AI curation and 1000x speed-to-data, ZL enables organizations to govern without copying a single file.



ZL Tech empowers enterprise GenAI with virtually governed and curated data sets