2025

インプレース
データ管理の台頭

ZL インプレースデータ管理ホワイトペーパー

ZLテクノロジーズ | WWW.ZLTI.COM

ZLTECH

エグゼクティブサマリー

大規模組織にとって最も喫緊の技術課題は、企業データを競争優位性につなげ、AIに活用することです。しかし、非構造化データの量が爆発的に増加するにつれ、組織は情報の管理とそこからの価値抽出において、ますます大きな課題に直面しています。データの複製とコピーの保存を基盤とする従来のアーカイブおよび分析モデルは、ストレージコストとガバナンスリスクを考慮すると、非構造化データの量に対応できません。

インプレースデータ管理は、企業の情報管理方法に根本的な変化をもたらします。ZLプラットフォームは、文書自体をコピーすることなく、あらゆる文書の「本質」を抽出することで、組織が非構造化データを検索、整理し、その真の価値を引き出すことを可能にします。同時に、プライバシー、コンプライアンス、eDiscovery、記録管理といった情報ガバナンス要件を満たすソースレベルでの管理を可能にします。

大規模な非構造化データ管理の課題

21世紀のデータ爆発により、非構造化データの管理における従来のパラダイムは、ほとんどの大企業で数十ペタバイト、数百ペタバイトにまで膨れ上がり、もはや通用しなくなっています。蓄積されるデータの80%以上がメッセージやファイルといった非構造化形式で存在するため、組織は法規制やコンプライアンスに関するリスクに過度に直面する一方で、真の価値を見出すことができません。人間によって人間のために作られた非構造化データに含まれるインテリジェンスは、機密性と同じくらい重要です。価値が高いがゆえに、管理が難しいという問題もあります。このデータには、従業員の意図、感情、そして動向が含まれており、AIのトレーニングにとって非常に戦略的な資産となります。

従来のアーカイブの壁

長年にわたり、法務、規制、ビジネスなど、様々な非構造化データを扱うアプリケーション向けに、多くのソリューションが登場してきました。アーカイブは増大する企業データを保存および保持するためのソリューションを提供し、エンタープライズコンテンツ管理（ECM）システムはビジネスレコードの管理に、データレイクは処理用データの保存に使用されます。これらのソリューションはそれぞれ、本質的に重複したデータが存在するため、追加のストレージコストが発生し、法的、規制的、サイバーセキュリティ上のリスクにさらされることになります。企業がそれぞれの目的のために新しいリポジトリを追加すると、管理すべきサイロがさらに増え、各サイロ間の通信が制限され、データ管理能力も異なります。その結果、プライバシー要件に準拠した個人データ管理など、複雑な機能の実行が事実上不可能になります。こうした機能は、他のガバナンスポリシーと緊密に連携させる必要があり、さらに今日の非構造化データの量を考えると、これらすべてのデータを（場合によっては複数回）コピーするコストは、現実的ではありません。

このようなアプリケーションの制約の結果、企業データ全体のうち管理されているのは比較的少なく、大部分は制御不能な状態にあります。組織がAIイニシアチブを推進していく中で、データのコピーや重複を避けながら、組織全体のあらゆる非構造化データからインテリジェンスを抽出できるよう管理する必要があります。

AI：サンドボックスに閉じ込められた

企業は、非構造化データの全容にアクセスできないため、企業データの完全で豊富なランドスケープ、つまり「ビーチ」ではなく、サンドボックスと呼ばれる小規模で不完全なデータセットでAIを訓練し始めています。これらのサンドボックスでは、AIシステムが知らないことを認識できず、与えられた情報からギャップを埋めるために幻覚を起こすため、信頼性が低下します。無関係な情報や不完全なデータセットではなくAIの真の価値を引き出すには、これらのモデルは、サンドボックスだけでなく、ビーチ全体から集められた最も関連性の高いデータにアクセスする必要があります。

削除の恐怖

企業は長らく、重要なデータ（規制記録、訴訟における潜在的な証拠、将来的に価値が証明される可能性のある情報など）を削除することへの懸念から、データ保持に偏りがちでした。しかし、ROT（冗長・不要・無駄）データの増加に伴い、過剰保持のリスクは増大しており、それらすべてを保管することに伴うコストも増加しています。特にデータのクラウドへの移行が進むにつれて、この傾向は顕著になります。また、過剰なデータ保持は、サイバーセキュリティ侵害のリスク、そして法的・規制上のリスクを大幅に高めます。

今日の企業は、規制、法律、ビジネス要件に準拠したポリシーに基づいて、時間の経過とともに「防御可能な」データ削除を可能にするソリューションを求めています。

データへのスピードの問題

企業は時間の経過とともにデータを整理する必要があるだけでなく、関連情報を即座に見つけて活用するためには、電子メール、ファイル共有、SharePointサイト、コラボレーション プラットフォームなど、すべての非構造化データ ソースからすべての単語をスクレイピングできる強力な検索アーキテクチャが必要です。

一部の企業は、Microsoft 365などのメールやコラボレーションプラットフォームのネイティブ機能に依存するようになりました。しかし、これらのシステムはユーザーの生産性と部門間のコラボレーションのために設計されています。その結果、単一のマスターインデックスではなく、数千もの個別のインデックスに依存するアーキテクチャが生まれました。つまり、検索には数千もの個別のインデックスを解析する必要があり、速度と信頼性が大幅に低下し、厳格なeDiscoveryやコンプライアンス要件を満たせないケースが多くあります。さらに、フルテキストインデックスを維持しないソリューションは、規制や企業レコードの定義が変更されたときに、データ全体を再処理する必要があります。

組織がMicrosoft 365に保存されている重要なデータを要求すると、情報の抽出は「スロットル」されます。つまり、一度に抽出できるデータ量に制限があるため、エクスポートに数時間、数日、場合によっては数週間かかることがあります。新しい抽出が開始されるたびに処理が停止すると、データへのタイムリーなアクセスが不可能になります。

データ量とガバナンス要件は、ネイティブプラットフォームの能力を上回っています。GenAIの登場により、組織が**企業全体のデータに反復的かつ進化的にアクセスし、管理する**ことが不可欠になっています。

情報ガバナンスとAIおよびアナリティクスとの融合

企業は大規模なAI導入において大きな課題に直面しています。公開データで訓練されたオープンソースのAIプラットフォームは、企業にとって関連性があり有用な出力を生成することができません。

運用管理は容易ではなく、セキュリティ上の懸念から、これらのプラットフォームは企業データを用いた学習もできません。そのため、組織は自社データを用いて独自のAIモデルを開発せざるを得ません。社内AIは非構造化コンテンツに大きく依存しますが、ターゲットを絞った関連性の高いデータが必要であり、ROTデータや機密データといった不要なデータからフィルタリングする必要があります。 今日、組織は、ガバナンス、リスク、コンプライアンスの観点からデータを適切に管理しながら、企業全体からAIに最も関連性の高いデータを見つけ出し、選別し、提供することに苦慮しています。

適切に管理されていないデータは、AIによる分析結果の信頼性を低下させ、リスクの増大につながります。現状では、AIが機密情報を取り込んだり、些細な情報や信頼性の低いコンテンツをフィルタリングしたりすることを防ぐためのガードレールはほとんどありません。一度AIをトレーニングしたら、それを解除することはできません。そうすることは、プールからインク一滴を取り除こうとするようなもので、 不可能です。 今日欠けているのは、eDiscovery、プライバシー、 レコード管理、規制遵守など、非構造化データに対する「ガバナンス・ガードレール」です。

従来のアーカイブで遭遇する障壁を打破するには、組織はデータ管理に対して根本的に異なるアプローチを必要とします。

ZLプラットフォームによるインプレース管理

インプレース管理：非構造化データの未来

これまで組織は、レコード管理、アーカイブ、電子情報開示用のリポジトリを作成するなど、さまざまなビジネス要件に合わせて非構造化データの約 5% を管理してきましたが、今日では非構造化データ環境全体を管理する必要があります。 残りの95%に対応するため、ZLプラットフォームはデータを「仮想的に」、つまり元のドキュメントをコピーせずにメッセージやファイルを管理できる、インプレースで管理するように構築されています。このプラットフォームは、メタデータやコンテンツを含むすべての文書のエッセンスを抽出してインデックス化し、 契約書などの高価値文書のみがアーカイブされます。これにより、関連性が高く、 管理されたデータを検索、整理し、AIリポジトリにフィードすることが可能になります。

単一の統合プラットフォームから、ユーザーはデータソースに対して、ドキュメントの分類、保持、レビュー、そして正当な削除 といったレコード管理機能を実行できます。手動および自動のデータタグ付けにより、ユーザーはドキュメントを分類し、規制要件に準拠したカスタムポリシーを適用できます。さらに、ユーザーは高価値ファイルをレコードとして宣言し、個人識別情報（PII）を修復し、ROTファイルに削除対象フラグを設定することができます。

インプレースデータ管理により、単一のプラットフォームから迅速な再分類とポリシー更新が可能になり、サイロ化が解消され、不整合が軽減されます。 これにより、 組織はストレージコストを削減し、法的リスクを軽減し、 運用を簡素化することが可能になります。

インプレースデータ管理により、単一の統合プラットフォームから完全な情報ガバナンスを実現

ホールビーチ：AIのトレーニングと電子証拠開示の実施

インプレースデータ管理により、企業は不完全なサンドボックスではなくエンタープライズデータ全体にアクセスできるようになり、より徹底したAIトレーニングとeDiscoveryの検索が可能になります。例えば、ある大手金融機関は、大量のメールやメッセージをAIで分析・精査し、規制違反やその他のビジネスリスクの有無を確認しようとしていました。しかし、当時のAI技術には限界があり、1日に処理できるメッセージは最大2万件にとどまっていました。2万件ものメッセージを収集する唯一の方法は、一度に1つのサンドボックスからサンプルを抽出し、そのサンプルに関連情報が含まれていることを期待することでした。

これらの制約により、2つの根本的な課題が生じました。第一に、分析に最も関連性の高いデータのみを検索、選別、配信する必要がありました。これは、企業全体に散在する非構造化データの無秩序な性質を考えると、特に困難な作業でした。第二に、分析対象として選択されたすべてのドキュメントが、ガバナンス、ライフサイクル管理、そして機密コンテンツのフィルタリングを確実に実施されていることを確認する必要がありました。

この金融機関はZLTechに依頼し、ZLプラットフォームを導入することでデータ管理と企業全体の検索を統合しました。ZLの活用により、同社はAI処理能力の限界内で関連コンテンツを特定・抽出し、エクスポートされたすべてのデータがガバナンス、プライバシー、コンプライアンス基準を満たしていることを保証しながら、効率的にデータ全体を網羅的に調査することが可能になりました。この成功事例は、大企業が責任を持って効率的にAIプログラムを導入できるよう、ZLがいかに支援できるかを浮き彫りにしています。

ZLのインプレースデータ管理は、eDiscoveryプロセスを効率化し、訴訟で優位に立つためにも活用できます。非構造化データの「ビーチ」全体を綿密に検索することで、従来のサンドボックスベースの検索よりも桁違いに多くの「決定的証拠」（訴訟に大きく有利な影響を与える証拠）を発見でき、eDiscoveryを武器にすることができます。これにより、企業はキーワード交渉において優位に立つことができ、訴訟の早期かつ有利な勝利または和解に貢献できます。組織が内部調査、監査、AIのトレーニングなどを行う場合でも、あらゆる非構造化企業データを検索し、必要な情報を迅速かつ安全に、そして確実に見つけることができます。

防御可能な削除によるキュレーション

企業が膨大な企業データへのアクセスを目指す中で、膨大な量のROTデータに遭遇します。これらのデータを分析に活用する前に、これらのデータはクリーンアップし、適切に削除する必要があります。各ガバナンス機能とポリシーが整合していることを確認し、データが削除された理由を示す監査証跡を維持する必要があります。ZLプラットフォームは、企業がデータをキュレーション、分類、廃棄できるよう支援することで、データの関連性を飛躍的に高めながら、ストレージコストを最大60%削減します。

実際には、防御可能な削除は、企業のデータ環境におけるクリーンアップに大幅な進歩をもたらします。例えば、米国に本社を置くあるグローバル銀行は、世界中のファイル共有リポジトリで数十億もの文書を管理していましたが、データの拡散とROTによるリスクとコストの増大に直面していました。

当初、社内でデータ整理に取り組んだものの、1年間でわずか100万ファイルしか削除されず、ほとんど進展がありませんでした。銀行は一元管理体制が整っていなかったため、地域間でデータ保持ポリシーに一貫性がなく、管理されていないファイルが膨大に蓄積されていました。

2023年、同行はZL Techのインプレースデータ管理を採用し、年間1億件のドキュメントのクリーンアップを実現しました。このインプレースガバナンスへの移行により、大幅なコスト削減、法的リスクとプライバシーリスクの軽減が実現し、AI活用や選択的アーカイブ化など、より広範なデータ変革への取り組みの基盤が築かれました。

データ取得のスピード

Microsoft 365などの大規模データリポジトリからインテリジェンスを抽出することで、組織は大量のデータをAIおよびアナリティクスリポジトリにエクスポートする際に発生する「スロットリング」を回避し、データ取得までの時間を1000倍以上も短縮できます。Microsoftは通常、エクスポート前にデータセットを確認する機能をユーザーに提供することに制限を設けており、エクスポート速度自体も、一括エクスポートにはしばしば非現実的な速度に制限されています。さらなる反復処理が必要な場合は、このサイクルを繰り返す必要があります。

ZLプラットフォームは、データが作成されると、あらゆる文書のエッセンスを抽出し、ほぼリアルタイムで検索、選別、AIリポジトリへの送信を可能にします。関連情報が特定されると、企業は必要に応じて元の文書を簡単に取得し、保全したり、アナリティクスやAIリポジトリでそのインテリジェンスを活用したりすることができます。

情報ガバナンスがAIとアナリティクスを解き放つ

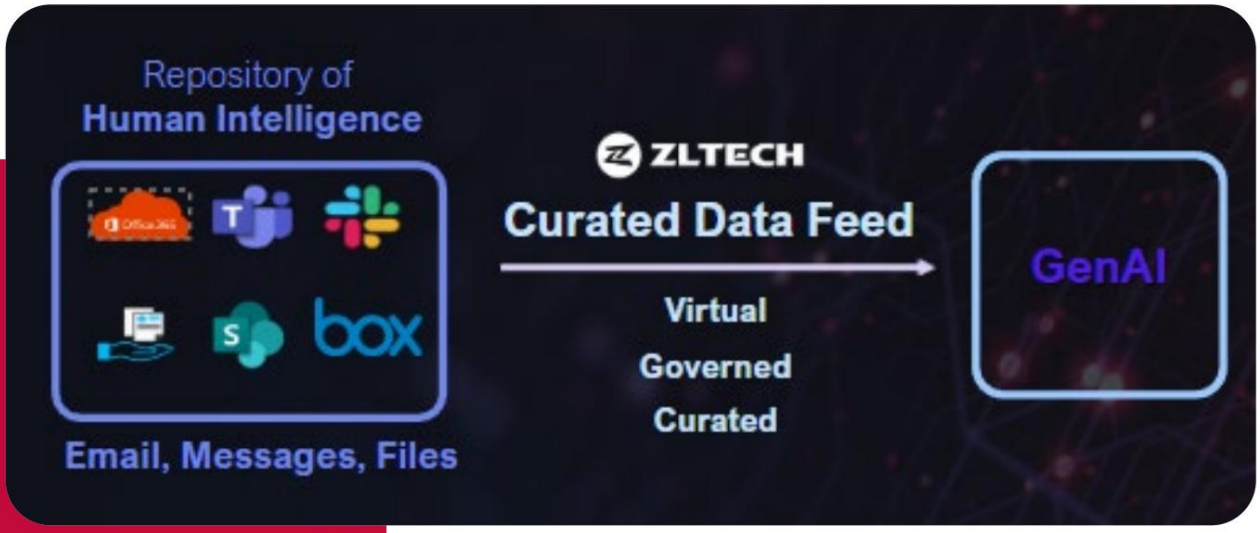
ZLプラットフォームは、膨大なデータに迅速かつ正確にアクセスできるだけでなく、あらゆる情報を貴重な企業インサイトへと変換する機能も備えています。従業員が作成・共有した情報を表面化させることで、企業のこれまで見えなかった一面を明らかにする力を持つことができます。非構造化データは、従業員とのあらゆるコミュニケーションの背後にある内容と意図を明らかにします。「頼れる」従業員は誰なのか？舞台裏でイニシアチブを主導しているのは誰なのか？

インプレースデータ管理は、企業のデータエコシステムを「仮想」データレイクに統合することで、社内AI向けにデータセットをキュレーションすることを可能にします。そこから、最も関連性の高い情報を検索、選別し、適切なAIシステムに配信することができます。関連性が高く完全なデータをAIにインプットすることで、幻覚を最小限に抑え、精度を向上させることができます。部門固有のアプリケーションは、それぞれのニーズに合わせてカスタマイズされたデータでトレーニングできるため、より正確な結果とより良いビジネス成果を実現できます。

ZLは、AIトレーニングセットから機密情報を除外することで、コンプライアンス、プライバシー、法的義務に関する露出リスクを軽減し、組織が自信を持って非構造化データを活用できるようにします。

結論

ZL Techのインプレースデータ管理は、従来のアーカイブの制約に縛られることなく、今日の企業のデータ量のニーズを満たす、情報ガバナンスへの革新的なアプローチを提供します。防御可能な削除から完全なエンタープライズ検索からAIキュレーション、1000倍のデータ速度まで、ZLは組織がファイルを1つもコピーせずに管理できるようにします。



ZL Techは、仮想的に管理・キュレーションされたデータセットで
エンタープライズGenAIを強化