



Knowledge Enrichment

Transforming Unstructured Data for Agentic AI
and Intelligent Solutions



Foreword

Organizations face a paradox: vast amounts of unstructured data with difficulty extracting value from it. As information management evolves from mere storage to strategic utilization, AI implementation success depends on high-quality, accessible data.

As a former CIO, I advocate a problem-first approach: defining business requirements before investing in solutions. This ensures AI systems deliver tangible value. Once objectives are clear, data enrichment provides context for both AI systems and human oversight.

Data quality is crucial—incomplete information leads to unreliable results regardless of AI sophistication. By contextualizing unstructured data, we transform it into knowledge that powers intelligent decision-making.

Tori Miller Liu, MBA, FASAE, CAE, CIP

President & CEO

Association for Intelligent Information Management (AIIM)



Executive summary

Unstructured data often remains trapped in silos or inconsistent formats, limiting efficiency, decision-making, and AI readiness.

Knowledge Enrichment transforms unstructured data into actionable insights through tools like extraction, metadata tagging and named entity recognition, which ensures content is ready for AI and automation systems.

As organizations advance toward more sophisticated AI implementations, including agentic AI systems that autonomously take actions to achieve goals, the quality and structure of data become increasingly critical.

Content enrichment and knowledge enrichment methodologies provide the foundation that these advanced systems require to function effectively. While basic content enrichment enhances discoverability, comprehensive knowledge enrichment creates the semantic understanding and contextual relationships that agentic AI needs to make intelligent, autonomous decisions.

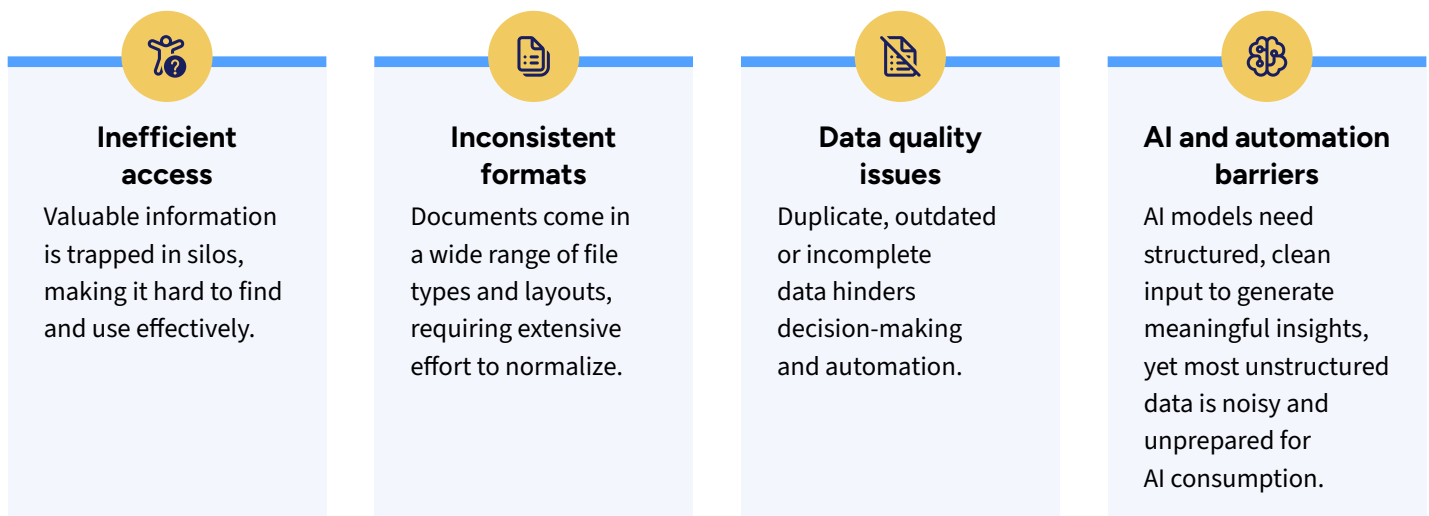
Organizations can improve data quality, decision-making, and operational scalability with these enrichment approaches. The techniques equip data engineers with tools to improve data quality, enable application developers to accelerate development and enhance functionality, and empower solution builders to innovate and create holistic solutions that can power next-generation AI capabilities.

The Unstructured Data Challenge

Unstructured data is everywhere — emails, contracts, reports, customer interactions, scanned documents and more. Unlike structured data, which fits neatly into databases, unstructured content is highly variable, making it difficult to organize, analyze and integrate into business processes.

The sheer volume is staggering: 80% of enterprise data is unstructured, and it is spread across multiple systems — on average, 21 different repositories within an organization, according to a [Hyland-commissioned Forrester study](#).

This fragmentation of critical content across repositories creates significant challenges, including:



Despite these challenges, unstructured data holds enormous potential. When processed correctly, it becomes a goldmine of insights, fueling AI-driven decision-making, automation and solution building.

The key to unlocking the potential of unstructured data is Knowledge Enrichment.



What is Knowledge Enrichment?

Knowledge Enrichment is an API-driven solution that transforms raw, unstructured content into meaningful, structured data that is ready for use in AI, automation, analytics and solution-building. This involves extracting key entities, applying metadata, linking contextual information, and ensuring content is machine-readable and actionable from the outset.

Knowledge Enrichment consists of two key components:

1

Data Curation is focused on structuring and normalizing unstructured data so it's clean and usable. This process is powered by [Document Filters](#), that enables the structuring of data from over 600 file formats while preserving the context of the original document.

2

Context Enrichment enhances unstructured data by linking it with relevant contextual information, which improves searchability, AI processing and decision-making.



How does Knowledge Enrichment change how you work?

Instead of relying on multiple postprocessing steps to clean and structure data, Knowledge Enrichment embeds intelligence early in the workflow, ensuring content is optimized for downstream applications as soon as it is ingested.



For example, traditional metadata for an image of a damaged house might include basic information like:

Image date: 10/10/2024

House Details:

- **Address:** 123 Main
- **Exterior:** brick, siding
- **Year Built:** 2015
- **Roof:** shingles
- **Stories:** 2
- **Home type:** single family



With knowledge enrichment, this can be expanded to include detailed analysis such as:

Damage Assessment

- **Location:** Main roof section, northwest quadrant
- **Area affected:** Multiple sections of roof shingles, potential decking exposure
- **Severity:** Moderate to severe wind/storm damage
- **Square footage affected:** Approximately 150-200 square feet

Visible damage includes

- Multiple areas of missing shingles exposing underlayment
- Several patches of displaced/lifted shingles
- One large area (approximately 4'x6') with exposed wooden decking
- Compromised water barrier in multiple locations
- Damaged flashing near roof peaks
- Ridge cap shingle damage visible along peaks
- Pattern of damage consistent with high wind event

Additional Observations

- Property is lakefront, potentially exposing roof to stronger wind conditions
- Surrounding houses visible in image appear to have intact roofs, suggesting localized damage
- Home appears to be relatively new construction
- Multiple elevation changes in roof structure may have contributed to wind damage patterns
- Current temporary repairs/tarpping may be needed to prevent water infiltration
- Professional assessment recommended for potential structural damage to roof decking



How Knowledge Enrichment works: A technical breakdown

The Knowledge Enrichment approach is powered by a combination of robust technologies and automation capabilities that ensure high-quality structured data output:

Data curation with proven tools



Document extraction technology extracts, normalizes and structures content from over 600 file formats, ensuring clean and consistent data that is ready for downstream applications while maintaining the document's logical structure. This allows organizations to seamlessly integrate data from diverse document types without losing its original meaning or intent.

Automated metadata tagging



Automated [metadata](#) tagging analyzes images to identify key elements, such as objects, scenes and text, then generates metadata to enhance searchability and AI model accuracy. This feature allows organizations to categorize and retrieve images more efficiently, improving content accessibility and relevance.

Data normalization and structuring



By converting unstructured text into standardized formats, Knowledge Enrichment makes content ready for machine learning, analytics and automation workflows. This process includes deduplication, format standardization and intelligent content segmentation.

Named entity recognition (NER)



NER technology identifies essential entities, such as people, organizations and locations from documents and assigns contextual meaning. This allows businesses to streamline categorization and automate intelligent workflows.

By automating these processes, Knowledge Enrichment:

- Reduces the reliance on manual data preparation
- Enhances data consistency
- Accelerates the time-to-insight for AI and analytics applications



Benefits of Knowledge Enrichment

From transforming unstructured content into actionable insights to enhancing AI performance and scalability, Knowledge Enrichment streamlines workflows and improves operational efficiency. Let's explore the key benefits of the solution, including its impact on data quality, cost reduction and decision-making.

Makes unstructured data actionable

By enriching data at the point of ingestion, Knowledge Enrichment turns raw documents into structured, AI-ready content. This eliminates the need for costly and time-consuming preprocessing steps. Additionally, by improving how data is organized and indexed, Knowledge Enrichment enhances accessibility, which allows businesses to quickly locate relevant information and take timely action. This ensures critical insights are readily available and reduces delays in decision-making while improving operational efficiency.

Improves AI and analytics performance

Clean, structured data leads to:

- Improved AI predictions
- Better decision-making
- Faster processing times

Organizations can trust that their insights are based on high-quality input.

Forrester Study: The Rise of Content Intelligence — A New Era of Innovation in ECM

In a recent survey conducted by Forrester Consulting and commissioned by Hyland, Forrester presented findings that underscore the importance of harnessing the power of enterprise content and unstructured data.

Explore their findings, which include insights on which content intelligence capabilities are ripe to flourish.

➔ [Download the report.](#)



Reduces operational costs

By embedding enrichment within the content workflow, businesses reduce the burden of manual data processing. This frees up resources and improves efficiency across teams and workflows.

Enhances scalability

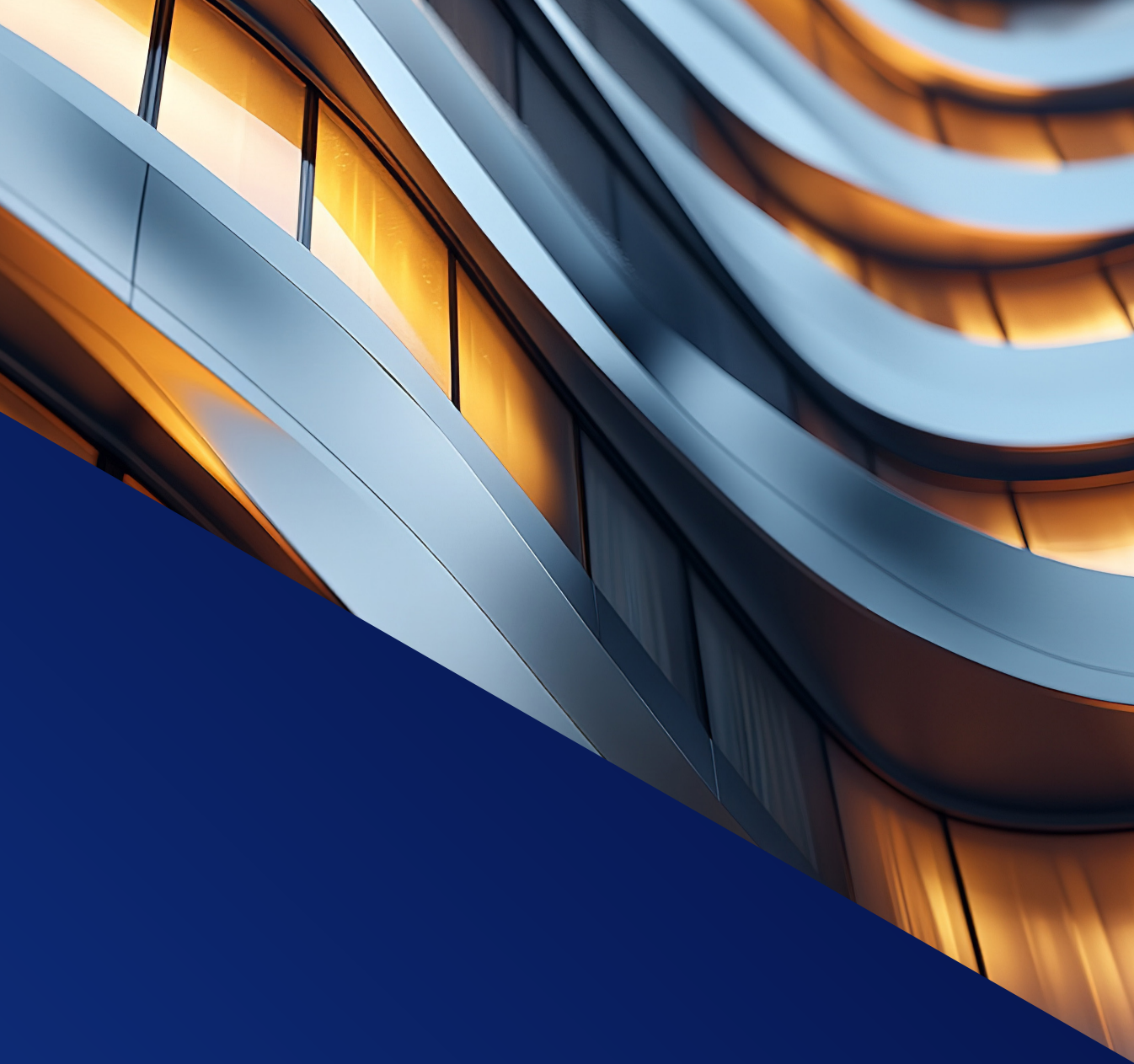
As organizations handle larger volumes of content, Knowledge Enrichment ensures they can process more data without added complexity, enabling seamless growth.

Improves data quality

High-quality data is the foundation of effective decision-making and AI-driven insights.

Knowledge Enrichment enhances data consistency, accuracy and completeness by structuring unstructured content, applying metadata and establishing meaningful relationships between disparate pieces of information. This ensures that organizations are working with the most reliable and relevant data possible, reducing errors and improving overall operational effectiveness.

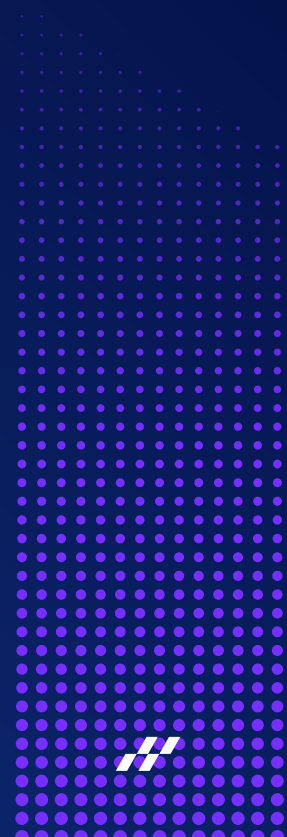




“

Unfortunately, poor content hygiene, like obsolete, untrusted, or duplicate information, is a common challenge. Sixty-two percent are taking steps to enrich or improve their content, making it more suitable for content intelligence. However, 52% say most of their content is not AI-ready.

The Rise of Content Intelligence: A New Era of Innovation in ECM,
Forrester, 2025



Real-world applications of Knowledge Enrichment

From improving product discoverability to streamlining claims processing and structuring patient records, Knowledge Enrichment empowers organizations to deliver better outcomes and scale seamlessly.

Let's explore real-world applications that showcase the value of Knowledge Enrichment in solving complex data challenges.

Retail: Automate metadata generation

A major retail company needed a solution to automate metadata generation and identify named entities across product catalogs. Knowledge Enrichment applied consistent metadata tagging and created contextualized descriptions. This enabled more accurate data about the documents and more efficient natural language search, as well as improved recommendation engines.

With Knowledge Enrichment:

- **Data Curation** extracted and structured product details across various document formats.
- **Context Enrichment** identified key attributes such as brands, specifications and categories.
- **Automated metadata tagging** enhanced search accuracy and product discoverability.

These enhancements led to more reliable product data, improved search relevance and better personalization in recommendations.



21

Average number of enterprise content repositories



65%

Say unstructured data is a largely untapped opportunity at their organization



89%

Organizations not maximizing AI use for content intelligence

— [The Rise of Content Intelligence: A New Era of Innovation in ECM, Forrester, 2025](#)



Insurance: Enhance claims processing

Consider an insurance company that processes thousands of claims daily. Traditionally, these claims contain a mix of structured forms and unstructured supporting documents, such as medical reports and photographs. Without Knowledge Enrichment, processing these claims would require extensive manual intervention, unnecessary slow approvals and higher operational costs.

By implementing Knowledge Enrichment:

- **Data Curation** extracts and structures content from diverse file types while preserving document context (consider pieces of an insurance claim that might include photos, documents, emails, etc.).
- **Context Enrichment** identifies key entities such as claimants, medical providers and vehicle details, linking them with relevant policy and claims data.
- **Image analysis** detects key aspects of submitted images, such as vehicle make and model, extent of damage, and potential inconsistencies, aiding in fraud detection and claims validation.

Applying Knowledge Enrichment results in faster claims processing, reduced manual workload and improved accuracy in fraud detection.

Healthcare: Structure patient records for enhanced insights

Healthcare providers deal with vast amounts of unstructured patient data, including physician notes, medical histories, prescriptions and test results. Extracting meaningful information from these records manually is time-consuming and prone to errors.

By implementing Knowledge Enrichment:

- **Context Enrichment** extracts key details such as diagnoses, medications, allergies and test results from diverse medical documents.
- **Data Curation** structures patient data into standardized formats, ensuring interoperability across electronic health record (EHR) systems.
- **Knowledge Enrichment** enhances decision-making by making critical patient information readily available for AI-driven analytics, predictive modeling and clinical decision support.

By transforming unstructured patient records into structured, actionable data, healthcare providers can improve operational efficiency, support better patient outcomes, and streamline compliance with regulatory requirements.

Knowledge Enrichment and Agentic AI

As AI advances, there is growing interest in agentic AI systems. These systems represent a significant advancement in AI capability:



What is Agentic AI?

Agentic AI refers to AI systems that are:

- Outcome-driven
- Autonomous
- Capable of memory, planning, and sensing
- Able to use tools
- Capable of reflection and learning

Unlike classical machine learning that recognizes patterns learned from data to predict or classify, or generative AI that creates new content in text, image, or other forms, agentic AI can take autonomous actions to achieve goals.



Enterprise Use Cases for Agentic AI

Agentic AI has numerous potential applications in enterprise contexts:

- Autonomous claims adjustment
- Workflow routing
- HR operations
- Customer success
- Fraud analysis



Challenges in Agentic AI Related to Unstructured Data

Agentic AI systems face several challenges when dealing with unstructured data:

- **Unstructured Data Integration** - Agentic AI systems may struggle to seamlessly integrate and process unstructured data sources, such as text, audio, and video, which are common in many industries.
- **Contextual Understanding** - Agentic AI agents may have difficulty comprehending the nuanced, contextual information often present in unstructured data, leading to potential misinterpretations or suboptimal decision-making.
- **Scalability and Performance** - The complexity of agentic AI systems may limit their ability to handle the high volumes and diverse formats of unstructured data, potentially impacting scalability and real-time performance.
- **Data Privacy and Security** - Agentic AI systems may raise additional concerns around data privacy and security, as they may have access to and process sensitive, unstructured information.
- **Transparency and Explainability** - The black-box nature of some agentic AI models can make it challenging to provide transparent and explainable insights to customers.

It's important to note that generic approaches will NOT work for agentic AI – these systems need DEEP understanding of vertical and specific workflows.



How Knowledge Enrichment Supports Agentic AI

Knowledge Enrichment provides the foundation for successful agentic AI implementation by:



Enriching and Contextualizing Data

Knowledge Enrichment feeds agents with the right context through:

- Metadata classification and categorization
- Summarization and analysis
- Content extraction, chunking, and vectorization
- Industry ontologies and knowledge graphs



Supporting Various Agent Types

Knowledge Enrichment can support different types of AI agents:

- Search agents (agents designed to locate and retrieve relevant information)
- ReAct agents (agents that combine reasoning and actions to complete complex tasks)
- Content management agents (agents designed to handle and organize enterprise content)



Unifying Content and Governance

Knowledge Enrichment helps establish consistent approaches to data preparation, allowing organizations to maintain control while enabling agentic AI to access and utilize information effectively.



Conclusion

With Knowledge Enrichment, businesses gain a competitive edge — turning unstructured data into a strategic asset that fuels smarter decision-making and innovation. As organizations move toward more advanced AI implementations, including agentic AI, the ability to effectively enrich and structure unstructured content will become increasingly critical to success.

Organizations implementing such systems should ask themselves these key questions:

- Have you identified the processes that you think AI will improve?
- Can you list the individual tasks that AI will automate within those processes?
- Have you identified and quality-controlled the pools of knowledge/data you will feed into the AI?
- Do you know which people and skills you will need to build and operate your AI solution?
- Have decision-makers within your organization agreed on which measurements you will use to define the AI solution's success?

By addressing these questions and implementing robust Knowledge Enrichment practices, organizations can better position themselves for success in an increasingly AI-driven business landscape.



Why Hyland and Knowledge Enrichment?

Hyland brings decades of expertise in [document processing](#) and [content intelligence](#), making us uniquely positioned to deliver high-quality data curation and content enrichment solutions. Our approach ensures:

- Data is structured and [AI-ready](#) from the start.
- Workflows are streamlined and scalable for future growth.
- AI-driven insights are faster, more accurate and cost-effective.
- Document Filters provides unparalleled support for over 600 file formats, ensuring data curation works seamlessly across a broad range of document types while preserving context.

With Hyland Knowledge Enrichment, businesses gain a competitive edge — turning unstructured data into a strategic asset that fuels smarter decision-making and innovation.

➤ Learn more about [Hyland Content Intelligence](#).



