

WHITE PAPER

# Content enrichment with semantic metadata

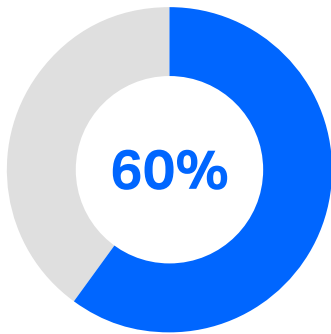
Bring structure to the information chaos caused by high volumes of unstructured content.

Assigning metadata to unstructured content has historically been a daunting, manual and time-consuming task. With artificial intelligence, enriching content with structured metadata can be automated and consistent. As a result, organizations reap significant benefits from having organized, findable content to increase productivity, support legal and compliance initiatives and empower data-driven decision-making.



## **Contents**

Executive summary	3
The growing challenge of unstructured content	3
Structured metadata is critical	4
Automation and metadata	4
The solution: Metadata enrichment for unstructured content	5
Value from content enrichment helps businesses	6
Finding information easily	6
Reducing risks and gaining control over information	6
Enhancing business processes	6
Improving customer experiences	6
Optimizing analytics with more data sources	6
Why don't organizations already have robust metadata?	7
Magellan Content Enrichment from OpenText	7



The volume of incoming content is expected to grow by 4.5 times over the next two years, with nearly 60 percent of that content being unstructured<sup>3</sup>

## Executive summary

Each day, the volume of business data continues to grow. In 2020 alone, 64.2 ZB of data was created or replicated, the majority of which is unstructured.<sup>1</sup> And the volume of incoming content is expected to grow by 4.5 times over the next two years, with nearly 60 percent of that content being unstructured.<sup>2</sup> Lacking organization and context, unstructured content is more difficult to retrieve, categorize and analyze. This hinders online collaboration, communication and information sharing, as well as discovery of content needed for legal and compliance purposes.

Artificial intelligence (AI) can be used to understand the meaning of unstructured content and the important elements within it—structured metadata. By using the metadata to give structure to unstructured content, organizations can proactively deliver content to improve the decision-making that drives productivity and processes.

For organizations ready to benefit from consistently and extensively enriching content with metadata through AI-based automation, the Magellan™ Content Enrichment solution from OpenText addresses the challenges of making unstructured content usable and findable at scale. It discovers existing metadata hidden within unstructured text and identifies additional metadata based on content and context. Magellan Content Enrichment automatically populates the semantic metadata into the content repository. Properly tagged and indexed content is ready for easy retrieval, workflows and cross-analysis with other data. This makes each document or piece of content more valuable at all levels across the organization.

Simply having this content isn't enough for organizations to benefit from the information that lies within it. To empower data-driven decisions and reduce the administrative burden of retrieving key information, forward-thinking organizations are turning to AI and content enrichment. This position paper offers a deeper understanding of content enrichment, including the process of classifying content with metadata to increase its value, along with its tangible business benefits.

1 IDC, DataSphere and StorageSphere forecasts. (2021)

2 AIIM, AIIM 2021 State of the Intelligent Information Management Industry: A Wake-Up Call for Organization Leaders. (2021)

3 ibid



## The growing challenge of unstructured content

Each day, the volume of business data continues to grow. In 2020 alone, 64.2 ZB of data was created or replicated, the majority of which is unstructured.<sup>4</sup> And the volume of incoming content is expected to grow by 4.5 times over the next two years, with nearly 60 percent of that content being unstructured.<sup>5</sup> Content enrichment allows users to understand the context and information locked inside content at scale. While data is among the most valuable assets that organizations have, unstructured data presents unique challenges.

Unstructured content, such as Microsoft® PowerPoint® presentations, Microsoft® Word documents and social media posts, is more difficult to retrieve, categorize and analyze than structured content. Its interoperability is also limited, meaning it cannot be used freely across multiple applications. This hinders online collaboration, communication and information sharing, which are all especially vital for modern workforces who may be remote or hybrid.

While unstructured content is made up of information that can be challenging for traditional applications to classify and define, it can contain informational data elements such as claim numbers, customer numbers and contract values, among other factual pieces of information. Today, natural language processing (NLP), a branch of artificial intelligence, is often used to understand the meaning of unstructured content, and the important elements within it. Those elements are called out as structured metadata. By using the metadata to give structure to unstructured content, organizations can proactively deliver content to improve the decision-making that drives productivity and processes.

## Structured metadata is critical

American technologist David Weinberger once said, “The cure to information overload is more information.” While this may seem counterintuitive at first, it makes sense in relation to assigning the right descriptive data to existing unstructured content. Known as metadata, this descriptive data is assigned to content and creates structural connections between pieces so that related information can be quickly retrieved.

Once metadata is in place, organizations can combine the information and insights mined from the unstructured content stored in ECM, CRM, ERP and other business systems to make more informed decisions. They can also connect content within business processes, so that all information is readily available at the right time to the people who need it.

For enterprises and public sector organizations, top priorities include decreasing organizational costs, reducing risks and improving efficiency. Fortunately, enriching unstructured content with metadata offers benefits that flow into other businesses processes. In fact, metadata is at the heart of true digital transformation. For instance, companies can build record management rules within enterprise content management systems that use this metadata to streamline workflows.

## Automation and metadata

The only practical way to enrich unstructured content with structured metadata is to use an automated process. There are far too many stored files for humans to sift through, and even if an organization had the time and resources for such an overwhelming task, there’s also the risk of human error to consider.

<sup>4</sup> IDC, DataSphere and StorageSphere forecasts. (2021)

<sup>5</sup> AIIM, AIIM 2021 State of the Intelligent Information Management Industry: A Wake-Up Call for Organization Leaders. (2021)



Tasking knowledge workers with assigning metadata is rarely successful and the results are often inconsistent. When forced, employees will either enter generic metadata with little value or bypass the content management system altogether. Inconsistency is also a risk, as different individuals may select different tags based on personal interpretation when presented with the same content.

Manual content tagging leads to employees wasting time searching for information and recreating information that already exists. Recent research found that the average knowledge worker spends almost 12 hours per week on these tasks.<sup>6</sup>

In addition, inefficiency, inattention and errors in meta-tagging can have serious financial consequences. For example, incorrectly handling data as important as healthcare records and insurance claims can lead to lawsuits and regulatory fines. Assigning metadata, such as policy numbers, status and patient names, through an efficient and reliable automated system is key to maintaining consistency, accuracy and timeliness.

While there are clear benefits from tapping into unstructured content, properly managing, maintaining and using it is arduous. Fortunately, as an alternative to manual tagging, AI-based automation is available to extract information quickly and consistently from unstructured content and make workflows more efficient and valuable. It all starts with content enrichment.

## **The solution: Metadata enrichment for unstructured content**

In the AI-automated process known as content enrichment, NLP is used to understand the meaning and the important elements within text. Enrichment tools can read through vast stores of data at rapid speeds, applying taxonomies to classify and tag content with rich, relevant metadata, such as categories, named entities and key concepts.

Content enrichment enables the indexing and tagging of content across entire enterprises, helping companies discover, understand and manage their content. The process is efficient, consistent and reliable. Businesses are reaping significant benefits from having this organized, findable content.

6 APQC, How KM Should Support the Business in 2022. (2022)





## Value from content enrichment helps businesses

The ways in which enriching content with metadata can help businesses are far-reaching. Here are just a few of the most noteworthy changes organizations see after adopting a content enrichment strategy.

### Finding information easily

Employees spend a significant portion of time simply searching for information. Assigning metadata to content allows employees to find data they need quicker and more efficiently. Metadata-driven filters make searches simpler and more targeted for greater accuracy and speed.

Once content is tagged, other applications (whether analytics or processed-based) can leverage information that's easily found and understood. In other words, the content that was once sitting largely unused in databases becomes ready for analysis or use in data-driven processes.

### Reducing risks and gaining control over information

Data has value to an organization, but also carries risk. Best practice is for an organization to not retain content longer than required for legitimate business purposes or as prescribed by law. In fact, possessing information means that it's discoverable in cases of legal disputes or subject to compliance requirements by the industry or legislation.

Metadata helps organization understand their content and properly classify and manage it, so they don't expose themselves to undue risk.

### Enhancing business processes

Content enrichment enables streamlined workflows by automating content delivery and triggering tasks based on the metadata within contracts or other key documents. For instance, invoices can be routed to the appropriate parties for approvals or review, resulting in fewer manual tasks. Organizations can also optimize case management with simplified organization, routing and retrieval of all case-related content.

### Improving customer experiences

Thanks to the proliferation of emails, customer chat transcripts and reviews, today's businesses have more opportunities than ever to learn from customer feedback. However, because of the sheer volume collected, many organizations fail to extract the immensely valuable insights that are hidden within them. With capabilities such as sentiment analysis, categorization and summarization, metadata content enrichment solutions allow businesses to quickly classify, respond to and act on customer feedback, empowering richer customer experiences.

### Optimizing analytics with more data sources

Content enriched with metadata allows organizations to mine stores of data for insights, patterns and relationships, enabling more comprehensive and analyses. By associating structured metadata with unstructured pieces of content, organizations enable the creation of reports, dashboards and analysis with new insights. The insights themselves can range from understanding the types of documents within a repository and when they were last accessed, to understanding the overall sentiment, emotion and intent displayed within communications.

## Why don't organizations already have robust metadata?

If businesses already have such immense stores of content that could be better leveraged, why isn't it already tagged and categorized with metadata? Until recently, a set of challenges has prevented companies from efficiently applying metadata to their content. For example, pre-existing information architectures have presented hurdles for managing data efficiently. Plus, organizations have, on average, five different content systems and repositories where unstructured content exists in silos and is often redundant.<sup>7</sup> Finally, as discussed earlier, manual processes are time-consuming and prone to errors.

As a result, assigning metadata to unstructured content has historically been a daunting, manual and time-consuming task. Yet, metadata is crucial for finding and reusing information from across all content repositories.

To learn more about how Magellan™ Content Enrichment can benefit your business, contact [OpenText AI & Analytics Services](#).

- [Read the Content Enrichment solution overview](#)
- [Learn about the OpenText Semantic Strategy Workshop](#)

## Magellan Content Enrichment from OpenText

For organizations ready to benefit from consistently and extensively enriching content with metadata through AI-based automation, the Magellan Content Enrichment solution addresses the challenges of making unstructured content usable and findable at scale. It discovers existing metadata hidden within unstructured text, including informational data elements, and by identifies additional metadata based on content and context. We call this semantic metadata. Magellan Content Enrichment automatically populates the semantic metadata into the content repository. Properly tagged and indexed content is ready for easy retrieval, workflows and cross-analysis with other data. As a result, each piece of content or document becomes more valuable at all levels across the organization.

With several readily available structured metadata taxonomies and knowledge models, Magellan Content Enrichment is well-suited for a broad range of industries and general business uses. For example, there are general tags for events, dates, geographic locations and names, as well as categories specific to life sciences, such as drugs and diseases, or business vocabularies that describe business functions like accounting, legal, engineering or manufacturing.

Businesses can also leverage custom options and pre-trained categorization models. Magellan can draw from more than 100 taxonomies to quickly start the process of building models to classify and tag content and can infer and update content inside OpenText™ Extended ECM Platform, OpenText™ Documentum™, OpenText™ eDOCS and other content repositories.

## About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: [opentext.com](https://opentext.com).

## Connect with us:

- [OpenText CEO Mark Barrenechea's blog](#)
- [Twitter](#) | [LinkedIn](#)

<sup>7</sup> AIIM, 10 things you need to know about digital transformation. (2020)