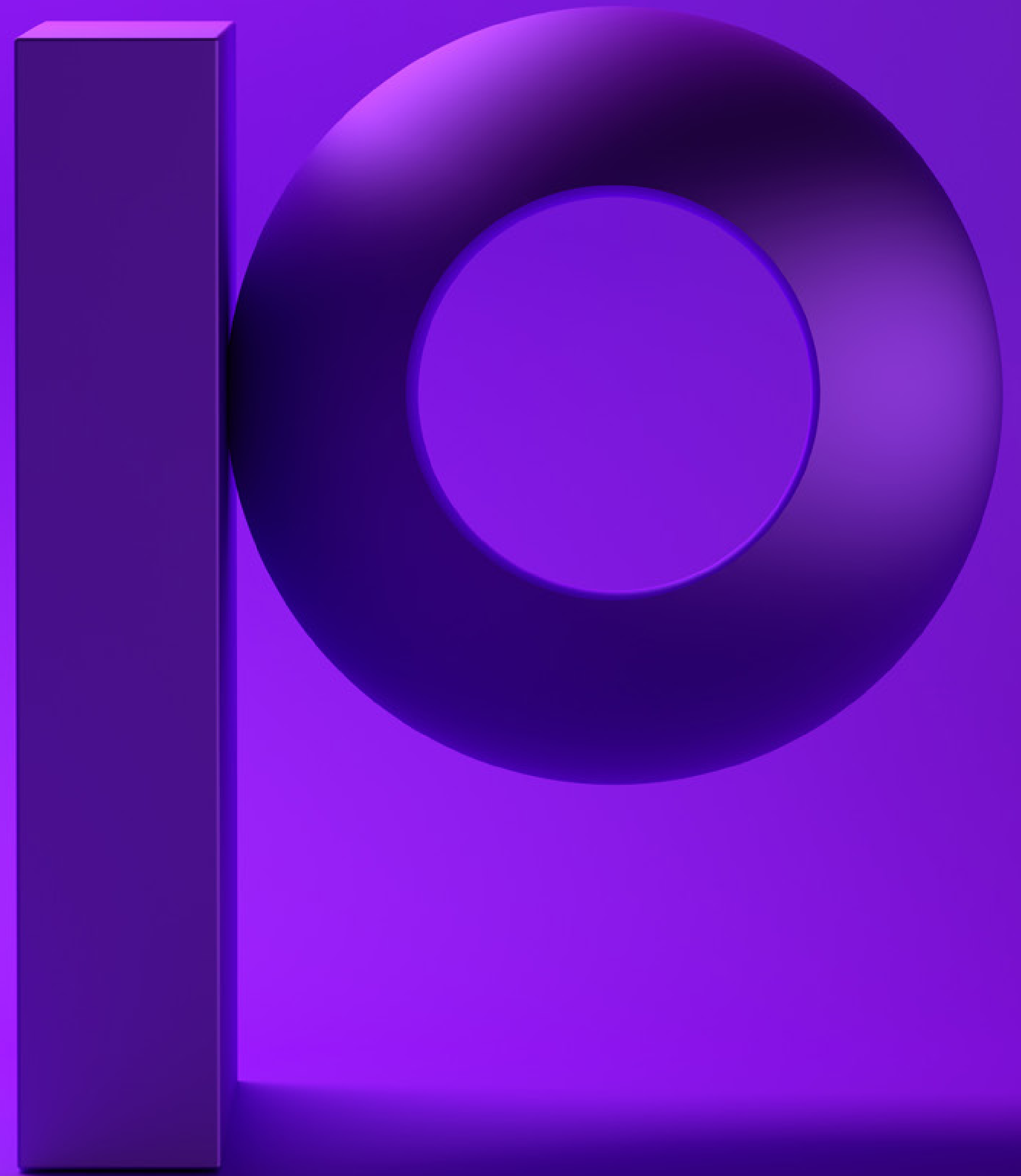


precisely

Best Practices for Using Unstructured Data



Introduction

PDFs, video files, images, customer communications, and multimedia content are all types of unstructured data organizations work with every day, proving to be a goldmine for those who can effectively unlock their use. While this data holds tremendous potential for insights, most organizations struggle to harness its power because extracting value from information that doesn't conform to predefined structures remains a significant challenge.

As AI and machine learning technologies mature, the ability to process and analyze unstructured data has become a critical differentiator, separating industry leaders from those left behind in an increasingly data-centric economy.



Unstructured Data vs. Structured Data

While structured data accounts for only about 20% of enterprise data, unstructured data comprises the remaining 80% - and continues to grow at an exponential rate. Unstructured data doesn't conform to a predefined data model and isn't organized in a traditional database format. Unlike structured data, which fits neatly into rows and columns with clear relationships, unstructured data exists in its native format without a specific schema. Examples include text documents, emails, social media posts, images, videos, audio files, web pages, sensor data, and PDF documents.

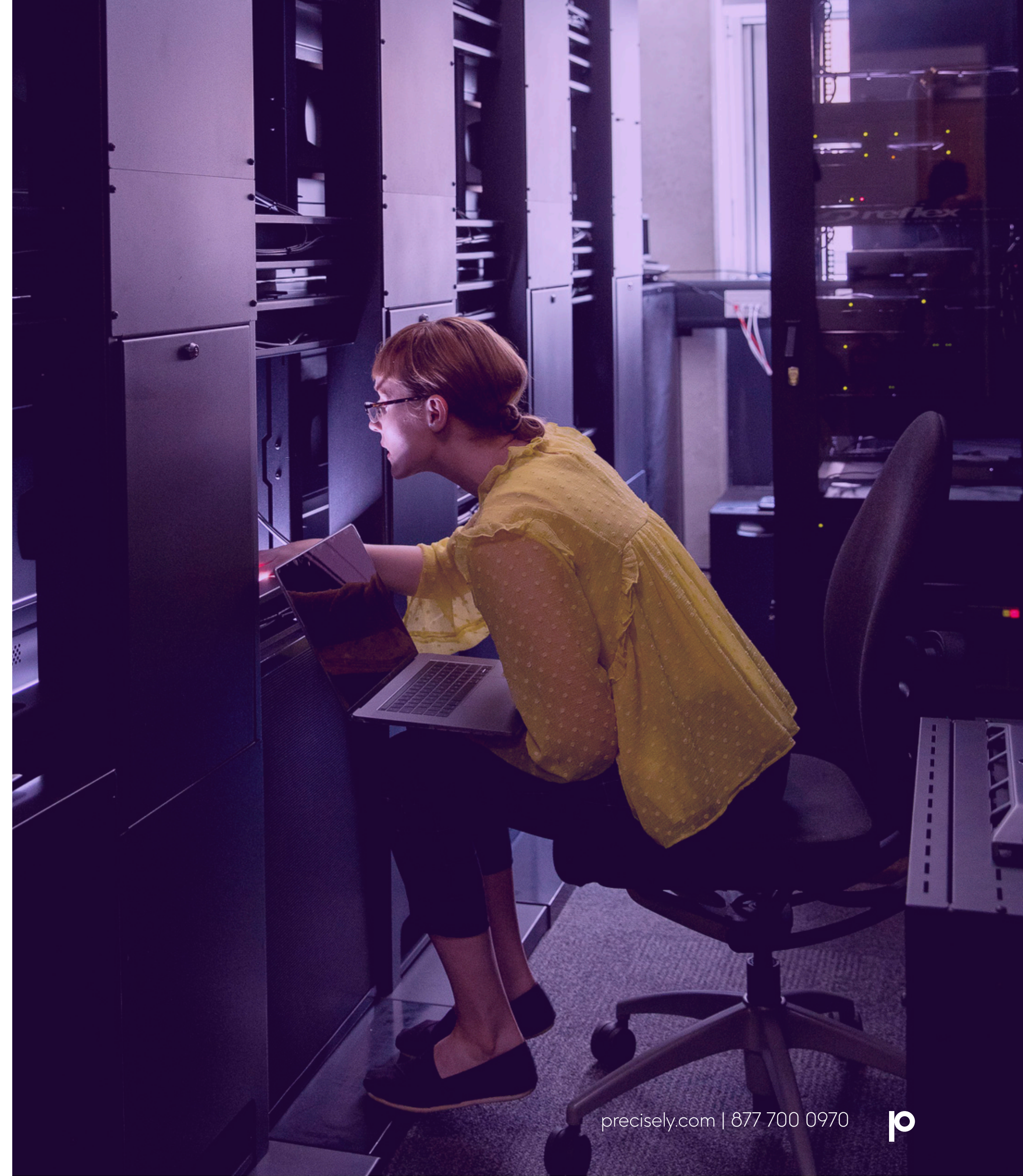
The distinction between finding value in structured versus unstructured data lies in how the data is organized and stored. Because structured data follows a rigid format, it's easily searchable and analyzable using traditional database queries. Unstructured data, however, requires specialized tools and techniques to extract meaningful insights because its value is often buried in free-form content.



The Growing Importance of Unstructured Data

AI systems excel at processing natural language, recognizing patterns in images, and extracting insights from complex, unorganized information sources. Large language models (LLMs) can analyze vast amounts of text to identify trends and actionable intelligence that would be impossible to uncover through traditional analysis methods. Computer vision algorithms can process images and videos to detect objects, faces, and behaviors, while natural language processing can extract meaning from documents, emails, and social media posts.

Unstructured data gets more useful as you collect more of it, since bigger datasets let you filter and spot patterns to create better targeted content. More data helps systems find small preferences and behaviors you can't see with less information, leading to better predictions and personal recommendations. Having lots of data supports testing different approaches, checking information across sources, and making improvements, creating a cycle where more data leads to better targeting and more effective, customized results that connect with what specific audiences want.



Use Cases for Unstructured Data

Organizations across industries are discovering innovative ways to extract value from unstructured data. In customer service, companies analyze support tickets, chat logs, and call transcriptions to identify common issues, improve response times, and enhance customer satisfaction. Sentiment analysis of social media posts and customer reviews provides real-time feedback on brand perception and product performance.

Financial services firms leverage unstructured data for fraud detection by analyzing transaction patterns, email communications, and the authenticity of documents. They also process news articles, analyst reports, and regulatory filings to inform investment decisions and assess market risks. Insurance companies use image analysis for claims processing, analyzing photos of damaged vehicles or property to expedite settlements.

Healthcare organizations extract insights from medical records, research papers, and clinical notes to improve patient outcomes and accelerate drug discovery. Retail companies analyze customer behavior through video surveillance, social media interactions, and product reviews to optimize store layouts, manage inventory effectively, and refine marketing strategies.

Manufacturing companies monitor equipment through sensor data, maintenance logs, and inspection reports to predict failures and optimize maintenance schedules. Legal firms process contracts, case files, and regulatory documents to identify risks, ensure compliance, and support litigation strategies.



Challenges in Managing Unstructured Data

The sheer volume and variety of unstructured data can overwhelm traditional data management systems. Storage costs escalate quickly, and processing requires specialized infrastructure and expertise that many organizations lack.

Quality and consistency issues plague unstructured data sources. Unlike structured data, which is characterized by built-in validation rules, unstructured data can contain errors, duplicates, inconsistencies, and irrelevant information. Extracting accurate insights requires sophisticated data cleaning and preprocessing techniques that demand considerable time and resources.

Privacy and security concerns are particularly acute with unstructured data, which often contains sensitive personal information, intellectual property, or confidential business data embedded within documents, images, or other forms of communication. Ensuring compliance with regulations such as GDPR, HIPAA, or industry-specific requirements becomes complex when dealing with diverse and unorganized data sources.



Most companies are still struggling with basic data governance challenges, so layering on additional privacy, security, and compliance requirements only compounds an already complex problem.

Integration challenges arise when attempting to combine insights from unstructured data with existing structured data systems. Different formats, schemas, and processing requirements can create silos that prevent comprehensive analysis. Additionally, the lack of standardized metadata makes it difficult to catalog, search, and govern unstructured data effectively.

Technical complexity represents another significant hurdle. Extracting meaningful insights from unstructured data requires advanced analytics capabilities, machine learning expertise, and specialized tools that many organizations don't possess internally. The rapid evolution of AI technologies also means that solutions can quickly become outdated, requiring continuous investment in new tools and training.



Organizational Preparedness Levels

Current research indicates that most organizations are in the early stages of utilizing unstructured data, with significant gaps between recognizing its value and implementing its capabilities. While approximately 85% of executives acknowledge that unstructured data contains valuable insights, only about 25% have implemented comprehensive strategies to systematically extract and analyze this information.

Many companies operate at a basic level, using simple text search tools or manual processes to handle unstructured data. These organizations typically struggle with data discovery, as they lack visibility into the unstructured data they possess and where it resides. They often rely on individual departments to manage their unstructured data sources, leading to inconsistent approaches and missed opportunities for cross-functional insights.

Some organizations have begun investing in specialized tools for specific use cases, such as document management systems or basic analytics platforms. However, they often lack integrated approaches that connect insights from unstructured data with broader business intelligence initiatives. These companies often face challenges scaling their initial successes across the entire organization.

Technically advanced organizations represent a small percentage of the market but demonstrate the potential value of comprehensive unstructured data strategies. These companies have invested in enterprise-wide platforms, developed specialized expertise, and established governance frameworks that enable consistent and scalable approaches to analyzing unstructured data.



Do You Need Unstructured Data?

Companies need to decide whether it's worth spending time and money to utilize their unstructured data, and this choice depends on the type of business they run. Some companies benefit significantly from analyzing documents such as customer emails, written reports, social media posts, and other text-based information. This is especially true for businesses in healthcare, banking, or law, where important details are often buried in documents and written notes. For these companies, identifying patterns in the written information can help them make informed decisions, enhance their services, and increase their revenue. Before diving into unstructured data projects, companies must clearly define their specific use cases and desired outcomes, as this understanding guides which data to collect, how to process it, and what success looks like.

However, many companies don't need to worry about this type of data analysis, and attempting to do so can waste a significant amount of time and money. Businesses that already have simple ways to track what matters most to them, or companies that don't have much extra money or staff, might not see enough benefit to make it worthwhile. Setting up the tools and hiring the right people to analyze written information is expensive and complicated. It also requires ongoing effort to maintain smooth operation and adhere to privacy regulations. Companies should carefully consider whether addressing their unorganized data will solve the real problems they have before investing their limited resources in it.



The Critical Role of Data Governance

Effective data governance serves as the foundation for successful unstructured data initiatives, providing the framework necessary to ensure quality, security, compliance, and the extraction of value. Without proper governance, organizations risk creating data swamps where valuable information becomes inaccessible or unreliable, rendering it unusable.

Data governance for unstructured data begins with establishing clear policies for data collection, storage, and retention. Organizations must define the types of unstructured data they collect, the duration of retention, and the circumstances under which it will be deleted. These policies must align with both legal requirements and business needs, as well as technical capabilities. To effectively ensure that governance is upheld, there are several elements to consider.

Metadata management becomes particularly crucial for effective governance of unstructured data. Rich metadata helps organizations understand the context, lineage, and business value of their unstructured data assets. This includes technical metadata about file formats and processing history, as well as business metadata about content, purpose, and relationships to other data sources. Classification and cataloging represent another critical piece, especially for governance components.

Organizations need to develop taxonomies that help identify and categorize different types of unstructured data based on content, sensitivity, business value, and regulatory requirements. Automated classification tools can help manage the scale of unstructured data; however, human oversight remains essential for ensuring accuracy and context.

Access controls and security measures must be tailored to the unique characteristics of unstructured data. Traditional database security models may not apply directly to documents, images, or multimedia files. Organizations must implement content-aware security measures that can identify and protect sensitive information, regardless of its format or location.

Quality management for unstructured data requires different approaches than structured data governance. Organizations must establish processes for validating data sources, identifying and handling duplicates, and ensuring that extracted insights are accurate and reliable. This includes implementing feedback loops that allow users to report quality issues and mechanisms for continuous improvement.



Conclusion

Successfully leveraging unstructured data requires a strategic approach that combines technological capability with robust governance frameworks. Organizations that invest in proper data governance, develop appropriate technical infrastructure, and build internal expertise will be best positioned to extract maximum value from their unstructured data assets. As AI technologies continue to advance and data volumes grow, the competitive advantage will increasingly belong to those who can effectively transform unstructured information into actionable business insights while maintaining security, compliance, and quality standards.





About Precisely

As a global leader in data integrity, Precisely ensures that your data is accurate, consistent, and contextual. Our portfolio, including the Precisely Data Integrity Suite, helps integrate your data, improve data quality, govern data usage, geocode and analyze location data, and enrich it with complementary datasets for confident business decisions. Over 12,000 organizations in more than 100 countries, including 93 of the Fortune 100, trust Precisely software, data, and strategy services to power AI, automation, and analytics initiatives. Learn more at www.precisely.com

www.precisely.com