

UNTAPPED VALUE: What Every Executive Needs to Know About Unstructured Data



Holly Muscolino
Group Vice President,
Content Strategies and the
Future of Work, IDC



Amy Machado
Research Manager,
Enterprise Content and Knowledge
Management Strategies, IDC



John Rydning
Research Vice President,
Global Datasphere, IDC



Dan Vesset
Group Vice President,
Analytics and Information Management, IDC

Table of Contents



CLICK ANY HEADING TO NAVIGATE
DIRECTLY TO THAT PAGE.

Introduction	3
In This White Paper	4
The Promise of AI and the Importance of Unstructured Data	5
A Unified Approach to Managing Unstructured Data	9
Underutilized, Undervalued, Underfunded	9
Complexity	11
Business Risks	12
Compliance Challenges	14
Productivity	14
Priorities	15
Benefits	16
Recommendations	17
Methodology	19
Survey Demographics	20
Appendix: Supplemental Data	21
About the IDC Analysts	22

Introduction

In 2022, 90% of the data generated by organizations was unstructured, and only 10% was structured. That year, organizations globally generated 57,280 exabytes of unstructured data — a volume that is expected to grow by 28% to over 73,000 exabytes in 2023. To put this in perspective, an exabyte is 1 million terabytes, or 1 billion gigabytes. Seventy-three thousand exabytes of unstructured data is equivalent to the amount of data in over 97 trillion sequenced human genomes; it's also equivalent to the amount of video streamed to 2.7 billion screens 24 hours per day for an entire year.

Unstructured data, or content, is mission critical for organizations. Unstructured data is integral to supply chains, where it is in purchase orders, product inventories, and import/export records. It is in sales agreements, marketing content, contracts, patents, movie scripts, patient treatment notes. It exists in financial earnings reports and employee performance records; it represents the core of design and engineering documents, product specifications, and product roadmaps. It constitutes the overwhelming bulk of human communication in the form of emails, meeting transcripts and notes, presentations, and instant messages. Perhaps most importantly, most of our knowledge is captured, curated, and shared in the form of unstructured data. Content is therefore essential to running a business, enabling organizations to embrace complexity, manage business risk, and increase productivity in the era of data and artificial intelligence (AI).

Understanding the role unstructured data plays within an organization is particularly significant considering generative AI (GenAI), which depends on large language models (LLMs) trained on massive volumes of text, images, audio, and video to identify patterns and drive productivity, creativity, and decision making at scale. Increasingly, leading organizations are recognizing that deploying GenAI depends on a mix of private and publicly available unstructured data.

With unstructured data, organizations are sitting on a veritable gold mine of information. Yet, according to IDC research, only half of an organization's unstructured data is analyzed to extract value, and only 58% of unstructured data is ever reused more than once after its initial use. If all data is to be considered as an asset — as echoed by many executives — then it must be treated as such. Imagine the reaction of investors or employees if a CEO or CFO is only willing to use half of a company's financial assets productively to generate a return.



of unstructured data **is ever reused more than once after its initial use.**

Yet that is the case with unstructured data, which is routinely wasted or neglected, exposing the organization to security risks and depriving it of opportunities to increase productivity, innovate with the latest AI functionality, and embrace data complexity to delight customers and contribute to all its stakeholders.

Most organizations are in this position because they are challenged by silos of unstructured data created, replicated, stored, and managed in myriad applications, tools, and systems. Many organizations lack a unified unstructured data security framework that balances governance with user access needs. Others rely on manual data classification, which is time consuming, affects ability to secure the data, and hinders ability to find relevant data fast.

IN THIS WHITE PAPER

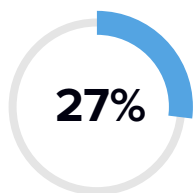
To better understand what differentiates leading organizations and their approach to unstructured data activation and utilization, IDC conducted a market research study, underwritten by Box, that surveyed over 400 business and IT decision makers, from large and mid-size organizations across industries and geographic regions. IDC analysts also conducted several in-depth interviews with Box customers, and relied on existing IDC syndicated research, such as IDC Global Datasphere, about data generation, replication, management, and utilization. Results of our research highlight the benefits of a unified approach to the management of unstructured data as a foundational capability in driving value in the age of AI.

The Promise of AI and the Importance of Unstructured Data

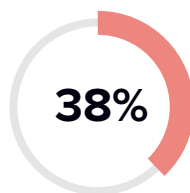
AI is not new, and many of us use it at work today, embedded into tools and applications to assist us in accomplishing day-to-day tasks. Classic AI or ML has seen broad adoption. However, a new branch of AI, called generative AI, has recently become available.

GenAI models trained on vast, mostly unstructured, publicly available data sets have caught the imagination of enterprise executives and consumers alike. This technology's promise to change how we work and interact with machines is already impacting the development of creative content, Natural Language Process (NLP)-based chatbots, text and video analysis, and myriad other use cases.

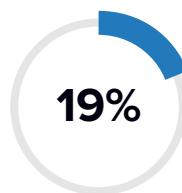
IN OUR STUDY...



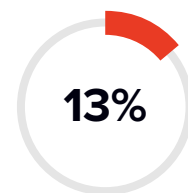
have already deployed generative AI broadly at their organization



have deployed generative AI in some areas



are actively exploring generative AI



are considering generative AI but haven't acted yet

Only 3% of our respondents are currently not considering deployment of the technology.

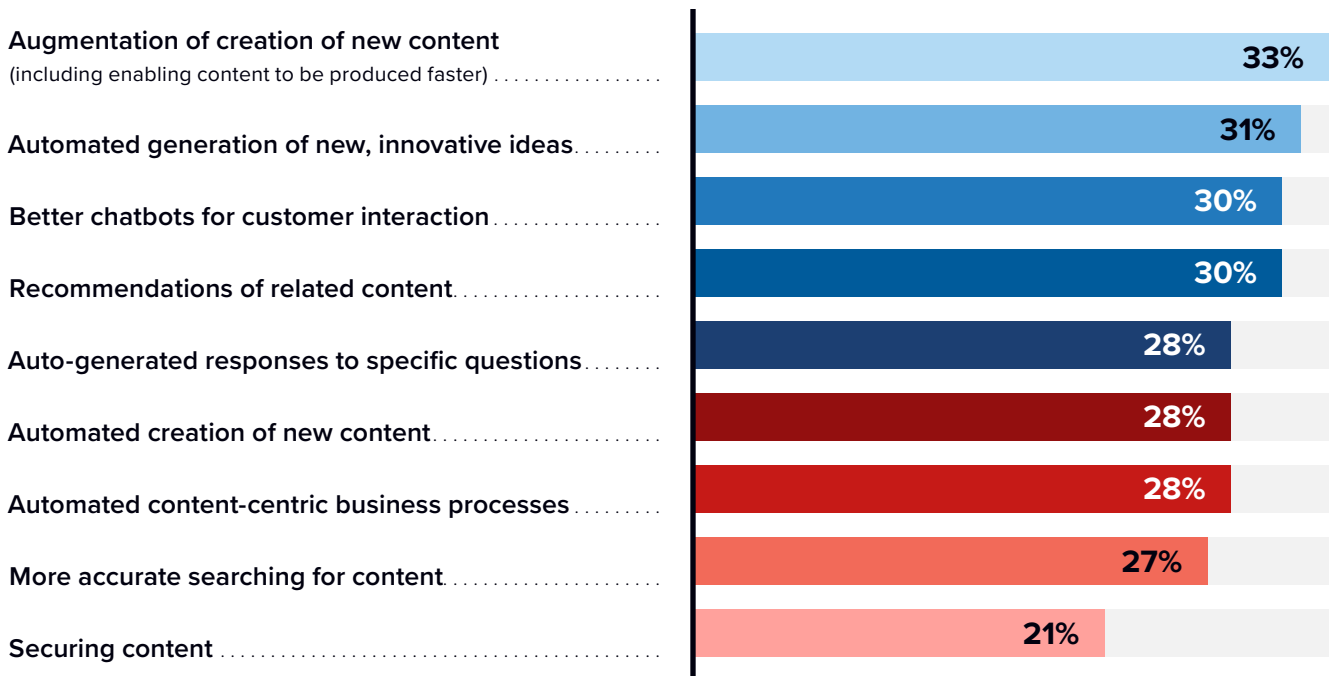
Organizations expect to gain several benefits from GenAI. **Figure 1** shows use cases with the greatest expected positive impact from GenAI. An underlying theme of these benefits is improved productivity — in business process optimization, creative content generation, human-computer interaction, knowledge creation and sharing, and other areas. This is especially important in a climate in which organizations are looking to do more with less. In IDC’s 2023 CEO survey, 61% of North American CEOs and 78% of European CEOs said there is an expectation to reduce the “run” part of their IT budgets to fund the new “build” initiatives in 2023. Organizations that don’t take advantage of this new opportunity risk losing ground to competitors.

FIGURE 1

Impact of Generative AI

Where do you expect to have the greatest positive impact from using generative AI?

(% of respondents)



n = 414; Source: IDC, *Unleashing the Opportunity for Unstructured Data Survey*, May 2023

Some specific use cases of GenAI:

- Identifying risky clauses in a contract
- Triggering a real-time business process via AI-extracted metadata
- Providing greater security at scale via AI-automation replacing manual interventions
- Quick search and validation of 3rd-party financial records to determine if a potential customer is a good fit for an offer
- Assessing the intent and tone of an email to determine the correct workflow — automated or human — for a better customer experience

As the excitement about the potential of GenAI has grown, so has the realization that its effectiveness depends on strategic management of an organization's unstructured data. Both GenAI and classic AI use cases depend on various AI models trained on a mix of internal, proprietary unstructured data and external open data. Whether that means training new models on long-form text, video, or image files, the expectation is that technology leaders develop a strategy to identify, classify, and unify all the available unstructured data in a secure manner and make it available to AI and data science experts without the risk of intellectual property leakage.

In fact, the number 1 most frequently cited roadblock to adoption of GenAI by our survey participants is concern about releasing the organization's proprietary content into the large language models of GenAI technology providers. This roadblock is followed by a lack of clarity about intellectual property rights around the content used as part of GenAI auto-generating new content and managing employee perceptions about automation of existing job functions.

The number 1 most frequently cited roadblock to adoption of GenAI by our survey participants is concern about releasing the organization's proprietary content into the large language models of GenAI technology providers.

As with any new technology, there are challenges and concerns. When asked about the biggest roadblocks to GenAI adoption, half (49%) of respondents in our study noted concerns about releasing their organization's proprietary content into the large language models of GenAI technology providers, and almost half (47%) of respondents cited lack of clarity about intellectual property rights around the content used to train large language models. The third most frequently cited roadblock to GenAI adoption is managing employee perceptions about automation of existing job functions (41%).

To address these concerns, technology providers with solutions that incorporate GenAI are doing the following:

- ✓ Committing to transparency about AI practices, technology, vendors, and data usage
- ✓ Providing full customer control of AI usage, data, and processes; customers may enable or disable the use of AI and decide whether AI should be applied to their unstructured data
- ✓ Ensuring that no LLMs are trained using customer data without explicit approval
- ✓ Providing users with a clear understanding of how their AI system functions and the rationale behind AI output to provide context
- ✓ Deploying trustworthy AI models that support accuracy, reliability, and safety of AI solutions
- ✓ Ensuring that AI systems adhere to the same controls and data access policies that determine access to unstructured data within their platform and system overall: This includes safeguarding customer data by implementing robust security protocols including encryption

A foundational capability for enabling GenAI or any other AI models for deriving value from unstructured data is a unified, governed, secure, and accessible platform for all unstructured data.

A Unified Approach to Managing Unstructured Data

We asked our study participants what the impact would be on their organization if they could implement software (or a cloud service) for managing all unstructured data in a unified, governed, secure, accessible platform.

Ninety-two percent told us that it would have a moderate to extremely positive impact on innovation and costs. Fifty-five percent of respondents expect a moderate to extremely positive impact on security, although another 25% expect a somewhat positive impact on security. Yet most organizations are challenged with unifying all their unstructured data.

Underutilized, Undervalued, Underfunded

Our ongoing research points to the fact that even though the volume and variety of unstructured data is vastly greater than that of structured data, spending on technology to utilize unstructured data is lower than that for structured data.

- Fifty-five percent say that less than half of all unstructured data is shared among employees or systems.
- Forty-one percent say that less than half of all unstructured data is reused (i.e., accessed more than once after the initial use).
- Twenty-two percent of unstructured data is unnecessarily replicated because organizations simply don't know what they have or how to find it.

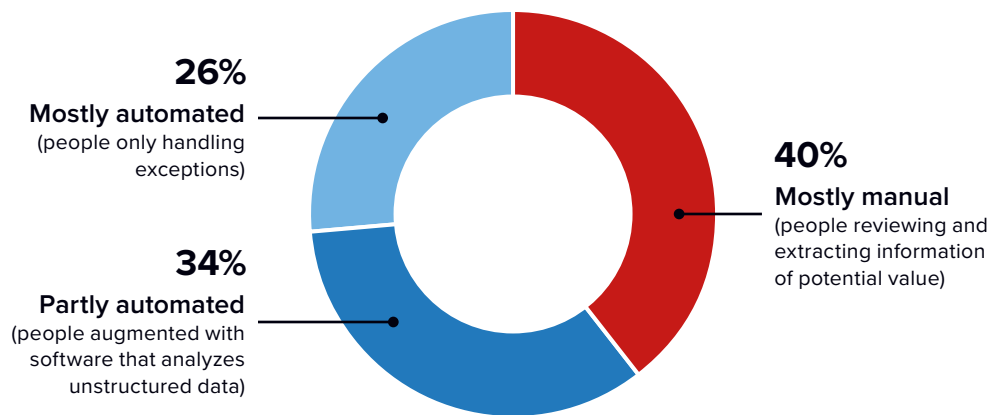
- Forty-six percent say that less than half of all unstructured data is being analyzed to extract value from it. Even then, when unstructured data is being analyzed, too much of it is happening manually as shown in **Figure 2**.

FIGURE 2

Methods for Analyzing Unstructured Data

What is your best estimate of the percent of unstructured data that is analyzed using the following methods?

(% of respondents)



n = 414; Source: IDC, *Unleashing the Opportunity for Unstructured Data Survey*, May 2023

Clearly, there is an opportunity for additional automation, which will not only improve productivity but will also enable organizations to leverage the value of unstructured data more fully.

These metrics suggest that too much of the potentially precious resource in the form of unstructured data is wasted or neglected. It doesn't generate a return on asset (ROA) and doesn't allow for downstream ROI.

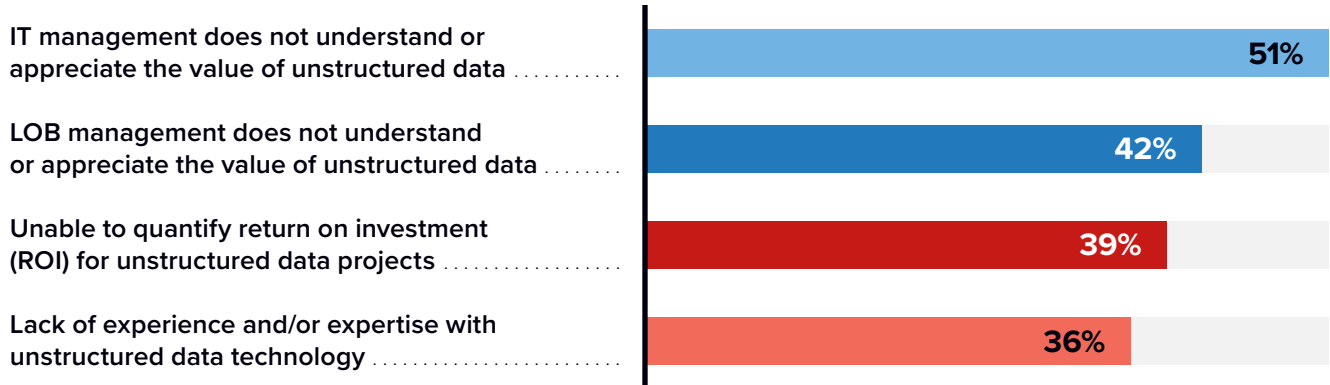
Yet only 44% of organizations say it is easy to justify funding for projects involving new or better utilization of unstructured data. The top inhibitors to justifying and obtaining funding for projects that utilize unstructured data are listed in **Figure 3** (next page).

FIGURE 3

Inhibitors to Fund Projects Leveraging Unstructured Data

What are the biggest roadblocks/inhibitors to justifying and obtaining funding for projects that utilize unstructured data? (Select all that apply)

(% of respondents)



n = 414; Source: IDC, *Unleashing the Opportunity for Unstructured Data Survey*, May 2023

Based on IDC’s ongoing research into technology spending, we estimate that only 40% of overall spending on ‘data’ technology is allocated to projects and initiatives with a focus on unstructured data. Increased funding is needed as organizations are both challenged by and seeking opportunities to derive benefits from unstructured data, which represents 90% of all data. And to grab the opportunity and value of unstructured data, companies must address four factors affecting and being affected by unstructured data: complexity, business risks, compliance challenges, and productivity.

Complexity

Half (50%) of our survey participants told us their company’s unstructured data is mostly or completely siloed. A quarter (25%) of organizations rate themselves as not being good at knowing and/or cataloging all the sources of unstructured data across the organization. That isn’t surprising when you consider that the typical employee regularly uses 37 software tools for their day-to-day work activities, and 70% of those are used to create, consume, or act on unstructured data. That introduces significant complexity into the work environment.

In fact, the top-ranked issue for organizations was the fact that the number of connections or relationships between various unstructured data sources and/or repositories is growing faster than employees’ ability to use/process/manage them. Twenty-eight percent of survey respondents noted that the fast growth of

connections or relationships between various unstructured data sources and/or repositories was a top challenge for them, while 40% find it difficult to integrate unstructured data technology with other technologies.

The speed and volume of data are also of concern. Twenty-five percent of survey respondents told us that data volume is growing faster than their ability to use, process, and/or manage it. Twenty-seven percent of respondents indicated that a top challenge was that unstructured data velocity was growing faster than their ability to handle it, and another 20% said that their top challenge was that unstructured data variety was growing too quickly. For example, one senior administrator at a large insurance company told us: *“If [our clients] get in an accident, they can record high-definition 4K video ... [the data] is only getting bigger and larger and more complex.”* But it’s important to remember that, even though all growing systems become more complex as they expand, a unified platform can simplify the management of unstructured data. In short, complexity doesn’t have to be complicated.

Business Risks

Application sprawl and the frequently resulting fragmentation of unstructured data create additional security and compliance risks for organizations. Both sprawl and fragmentation result in more potential attack surfaces that need to be secured — often with diverse sets of identity and authentication models and different administrative features, introducing more potential points of failure.

Our study participants indicated that their average annual costs of data security breaches were \$3.8 million. IDC estimates that this will be nearly \$4.5 million in 2023. However, organizations with more fragmented unstructured data approaches pay a heavier price for security breaches. We measured such fragmentation based on the most frequent method of unstructured data sharing — one being the use of point-to-point via email, chats, thumb drives, FTP, etc., and the other being via centralized sharing, content collaboration, or content management platforms. Greater fragmentation led to doubling of annual costs of security breaches (\$4.5 million vs \$2.2 million).

Organizations with more fragmented unstructured data approaches pay a heavier price for security breaches. Greater fragmentation led to doubling of annual costs of security breaches (\$4.5 million vs \$2.2 million).

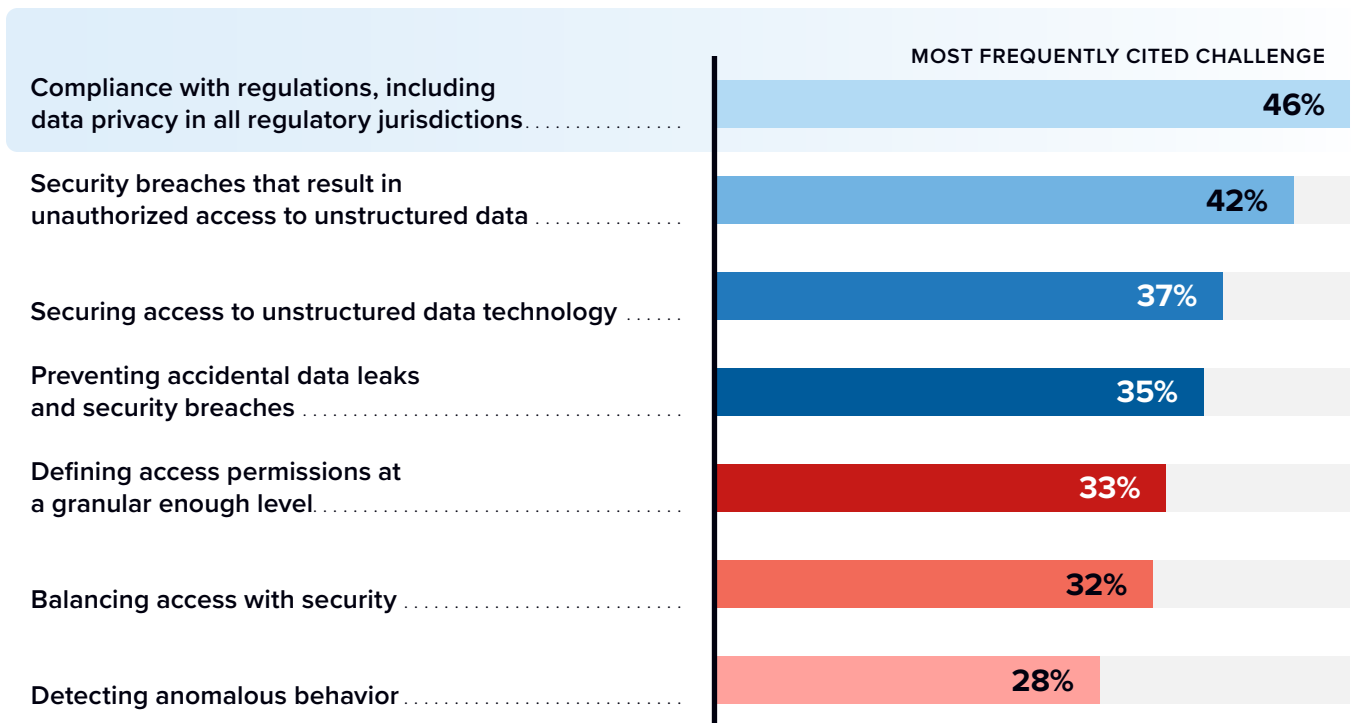
Organizations face a number of security challenges related to unstructured data, including unauthorized access to unstructured data, accidental data leaks, balancing access with security, and detecting anomalous behavior, as shown in **Figure 4**.

FIGURE 4

Security and Compliance Challenges

Which statements best describe your organization’s security challenges related to unstructured data?

(% of respondents)



n = 414; Source: IDC, *Unleashing the Opportunity for Unstructured Data Survey*, May 2023



51% of respondents reported non-compliance with data regulations in the past 12 months, with an average total cost of \$1.03 million.

Compliance Challenges

The most frequently cited challenge, as shown in **Figure 4** (previous page), is compliance with regulations. This challenge comes as no surprise for organizations dealing with content fragmentation across many apps in multiple geographies, and is likely to continue as new data and AI regulations are enacted across jurisdictions globally.

We asked our study participants how confident they were that their organization is retaining and/or disposing of unstructured data and, if necessary, making it available based on all current internal policies or external regulatory requirements. Almost half (46%) of respondents noted that compliance with regulations, including data privacy in all regulatory jurisdictions, is a top challenge for their organization. Over half (51%) reported non-compliance with data regulations in the past 12 months, with an average total cost of \$1.03 million. In addition to the cost associated with fines, non-compliance opens the door to brand and reputational risk as well as competitive risks for leakage of intellectual property in the form of unstructured data.

Our study also exposes a divide in confidence about internal and external compliance around retaining, disposing, and/or sharing their unstructured data. While 73% of companies are “somewhat” or “very confident” about complying with their own internal regulations, that assurance slips when it comes to external regulations. Only 59% have that same level of confidence about meeting external regulatory requirements. Of course, fragmentation of unstructured data increases the risk of non-compliance with industry and government regulations such as ITAR, GxP, GDPR, or HIPAA.

Productivity

Many of the shortcomings and challenges already highlighted in this paper have the dual issue of lowering employee productivity. IT staff are forced to spend more time securing unstructured data assets, or more time integrating siloed systems or addressing inefficiencies in collaborating with business staff or data scientists.

Business staff are hampered by their ability to either find relevant information or have it exposed to them within other applications in the flow of work. Inability to find the right information, at the right time, in turn leads to unnecessary duplication of effort in creating new unstructured data assets.

Many inefficiencies stem from factors noted previously: application sprawl and fragmentation of unstructured data and the applications that manage that data. We found that less than half (47%) of unstructured data is shared via centralized sharing, content collaboration, or content management tools, while 53% of unstructured data is shared point-to-point or person-to-person via channels such as email, chats, thumb drives, FTP, or shared drives. Sharing unstructured data via point-to-point or person-to-person methods inhibits productivity, as it makes it difficult to find that data and/or identify the correct version. It also introduces the potential data loss risk of unmanaged storage that can be lost or accessed by unauthorized individuals.

Importantly, fragmentation also inhibits knowledge sharing between employees and external stakeholders such as partners and suppliers. A critical issue for organizations is where either knowledge worker churn is high, or a significant portion of the employee base is expected to retire in the foreseeable future. However, knowledge classification and sharing is also a foundational capability for more productive innovation, employee upskilling, and improving the overall level of enterprise intelligence.

Despite the various challenges, the potential to derive value from unstructured data — whether using GenAI, classic AI, or enterprise applications — helps set corporate priorities. That means IT leaders need a content strategy, recognizing “the content problem” and the benefits that come with investments to address those challenges. A successful content strategy utilizes, values, and funds unstructured data projects.

Priorities

The funding process for unstructured data projects is improving: 40% of organizations expect funding of unstructured data projects to become easier over the next 1–3 years, while 40% expect the funding process to remain the same. It certainly seems that is one of the benefits of the popularization of GenAI.

ORGANIZATIONS’ TOP 3 PRIORITIES WHEN IT COMES TO BUSINESS DATA IN 2023:



1
Compliance and data privacy



2
Decision making that incorporates insights from unstructured data



3
Data security

ORGANIZATIONS' TOP 3 PRIORITIES WHEN IT COMES TO BUSINESS DATA WORKFLOWS IN 2023:



1 Knowledge retention/management



2 Contextualizing data for uses across the organization



3 Ensuring faster data availability

Benefits

We asked respondents how they would rate their organization on its ability to know and/or catalog all the sources of unstructured data generated across business units or departments. In other words, how good are they at knowing what types of unstructured data they have, what are its use cases, and where it is located? Forty-three percent rated their organization as “very good” or “excellent” in this regard, while 32% rated themselves as “good” and 25% said that they were “fair” or “poor”.

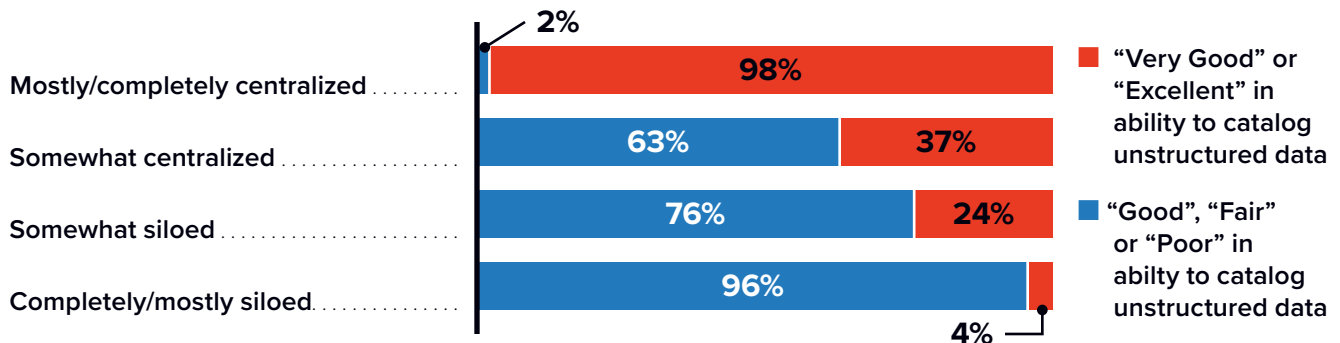
Of those 43% who say they have “very good” or “excellent” knowledge of their data, almost all have mostly or completely centralized their data. On the other hand, most organizations whose unstructured data is completely or mostly siloed are challenged in finding the resources they need (see **Figure 5**).

FIGURE 5

Centralizing Unstructured Data Significantly Improves Ability to Gain Value

To what extent is unstructured data in the tools, applications, and systems you selected stored and managed centrally (centralized) vs being stored and managed through each individual tool or application (siloed)? How would you rate your organization on its ability to know and/or catalog all the sources of unstructured data generated across business units or departments?

(% of respondents)



n = 414; Source: IDC, *Unleashing the Opportunity for Unstructured Data Survey*, May 2023
 For an accessible version of the data in this figure, see [Figure 5 Supplemental Data](#) in the Appendix.

In other words, having data that is mostly or completely centralized goes a long way in helping organizations know what they have and where it can be found. Data silos hurt the ability to leverage data for business insights, knowledge, and improved employee and/or customer experience. They can make non-compliance and security breaches more likely.

Centralization also offers the advantage of easier access to a large amount of data in one location for training large language models for generative AI. In short, those organizations without a unified content strategy are likely to be left behind.

We also assessed the extent to which utilizing unstructured data improved the following capabilities of the respondent's organization over the past 12 months.

The top 6 most frequently mentioned outcomes with greatest benefits included:

1. **Customer satisfaction, engagement, and retention**
2. **Data governance**
3. **Compliance with regulations**
4. **Innovation and/or creativity**
5. **Employee productivity**
6. **Collaboration**

These business benefits come from valuing — and then leveraging — unstructured data, with an investment in a centralized, secure, and accessible platform.

Recommendations

To not only remain competitive, but to thrive in the era of AI, organizations must treat their data as an asset. This is already largely the case for structured data held in databases and data warehouses. Yet treatment of unstructured data remains substandard to that afforded to structured data.

Our research showed the many challenges organizations face with siloed and highly varied unstructured data. We also found that organizations are interested in investing more in unstructured data initiatives as they realize the need for such data in training GenAI models. They recognize the risks and costs involved with not investing more in technology, skills, and processes related to deriving value from unstructured data.



of tech spend is allocated towards **unstructured data initiatives.**



of tech spend is allocated towards **structured data initiatives.**

Recent advances in platforms for unifying unstructured data to facilitate creative and business processes, compliance with regulations, employee productivity, and product or service innovation will make it easier for organizations to mine unstructured data for greater business value.

ORGANIZATIONAL LEADERS IN BUSINESS AND IT SHOULD CONSIDER THE FOLLOWING:

- 1** Take stock of your organization's unstructured data as well as processes that rely on it.
- 2** Evaluate the latest technology platforms to unify unstructured data.
- 3** When considering such platforms, assess platform scalability, performance, manageability, interoperability, and security.
- 4** Deploy a data classification scheme to support unstructured data access and utilization.
- 5** Assess unified data platform technology vendors' current AI offerings and their roadmaps in the rapidly changing AI technology.
- 6** Continuously experiment with GenAI to identify productive use of this new technology.
- 7** Initiate or expand a data literacy program to facilitate better interaction of employees with AI-infused data solutions.
- 8** Articulate the value of the human-in-the-loop approach to leveraging GenAI as a tool to augment employees by automating repeatable steps in their workflow rather than replacing human expertise, judgment, and interpersonal communications skills.
- 9** Utilize GenAI to augment the work of the compliance team to understand the impact of unstructured data on compliance with regulations.
- 10** Identify opportunities for integrating unstructured and structured data to optimize processes and improve data-driven decision making.

These steps will enable organizations to unlock the untapped value within their unstructured data.

Methodology

The viewpoints, analysis, and recommendations presented in this white paper are based on IDC's market research into the use and management of unstructured data. Part of the research included a survey of 414 business and IT decision makers, from large and mid-size organizations across industries and geographic regions, who are involved with ensuring access and/or use and management of unstructured data within their organizations. In addition to the survey, IDC conducted several in-depth interviews with Box customers, and relied on existing IDC syndicated research, such as IDC Global Datasphere, about data generation, replication, management, and utilization.

IDC researchers defined unstructured data as data that is not organized in a predefined manner and does not conform to a semantic database model that facilitates addressing and/or analyzing the data.

Unstructured data includes:

- **Long-form documents** (e.g., manuscripts, proposals, presentations, spreadsheets, memos, and other textual documents, such as statements/invoices and forms)
- **Short-form text** (e.g., emails, instant messages or texts, social streams)
- **Captured rich media** (e.g., surveillance photos/videos/audio, meeting videos, or video captured as part of operational processes)
- **Produced rich media content** (e.g., video, audio, images for creative, entertainment, marketing, sales purposes)

On the other hand, structured data is any highly organized data usually stored in a database, which conforms to a data model and where data is generally tabular with columns and rows. Structured data includes sensor and IoT data. Structured data may include transactional or operational data such as sales transactions, financial accounts, customer lists, or employee records.

See next page for the primary survey demographics.

SURVEY DEMOGRAPHICS

COMPANY SIZE



Mid-market
501–2,000
employees

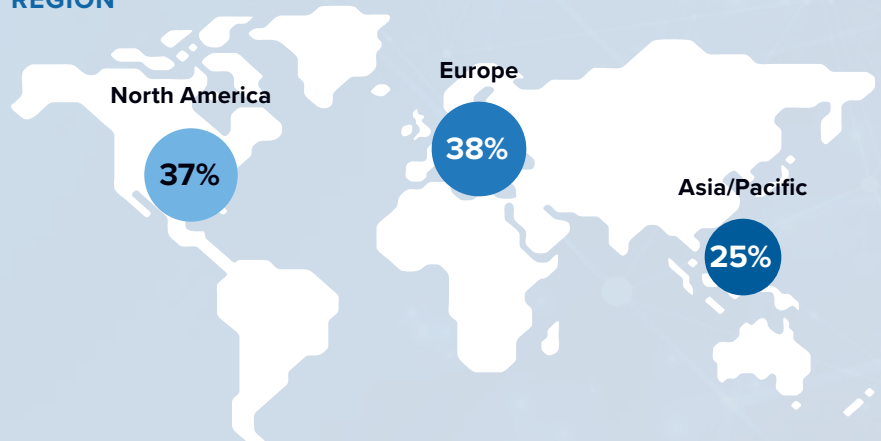
51%



Enterprise
2,001+
employees

49%

REGION



INDUSTRIES



Financial
Services

15%



Government

15%



Retail &
Consumer
Goods

15%



Media,
Entertainment,
Marketing,
Advertising

15%



Life
Sciences

15%



Other
Industries

25%

ROLE



Line of
business

26%



IT

74%

Appendix: Supplemental Data

This appendix provides an accessible version of the data for any complex figures in the document. Click “Return to original figure” to get back to the original data figure.

FIGURE 5 SUPPLEMENTAL DATA

Centralizing Unstructured Data Significantly Improves Ability to Gain Value

	“Very Good” or “Excellent” in Ability to Catalog Unstructured Data	“Good”, “Fair” or “Poor” in Ability to Catalog Unstructured Data
Mostly/completely centralized	98%	2%
Somewhat centralized	37%	63%
Somewhat siloed	24%	76%
Completely/mostly siloed	4%	96%

n = 414; Source: IDC, *Unleashing the Opportunity for Unstructured Data Survey*, May 2023

[Return to original figure](#)

About the IDC Analysts



Holly Muscolino

Group Vice President, Content Strategies and the Future of Work, IDC

Holly Muscolino is the Group Vice President, Content Strategies and the Future of Work, responsible for research related to innovation and transformation in content solutions, including intelligent document processing, esignature, imaging and printing, and other content workflow services. Holly's core coverage also includes work transformation, technology and digital skills research, and the role of technology in driving the Future of Work.

[More about Holly Muscolino](#)



Amy Machado

Research Manager, Enterprise Content and Knowledge Management Strategies

Amy Machado is a research manager for IDC's Enterprise Content and Knowledge Management Strategies coverage. She also collaborates on Intelligent Document Processing research. Her advisory program examines the fundamental redesign of content services and related use cases that are transformed by digitization and the application of innovative technologies. Driven by the increasing movement to the cloud and the demand for using data to make business decisions, these new technologies create many opportunities for automation and insights for what has traditionally been undefined and unstructured data. Amy covers software for customer communications management, enterprise content management, and content sharing and collaboration. She uses primary research to size, forecast, and segment these enterprise content and knowledge management tools and software markets.

[More about Amy Machado](#)



John Rydning

Research Vice President, Global Datasphere, IDC

John Rydning is responsible for the Global DataSphere forecast, which measures the amount of data created by year and also for the Global StorageSphere forecast, which is a measure of the installed base of storage capacity worldwide and the amount of data stored in any given year. Additionally, John leads insightful research that explores key trends, use cases, technologies, and other factors shaping both the Global DataSphere and StorageSphere.

[More about John Rydning](#)



Dan Vesset

Group Vice President, Analytics and Information Management, IDC

Dan Vesset is group vice president of IDC's Analytics and Information Management market research and advisory practice where he leads a group of analysts covering all aspects of structured data and unstructured content processing, integration, management, governance, analysis, visualization, and monetization. Dan also leads IDC's global Big Data and Analytics research pillar.

[More about Dan Vesset](#)

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. This IDC material is licensed for external use and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.



IDC Research, Inc.
140 Kendrick Street, Building B, Needham, MA 02494, USA
T +1 508 872 8200

 @idc

 @idc

[idc.com](https://www.idc.com)

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.

©2023 IDC. Reproduction is forbidden unless authorized. All rights reserved. [CCPA](#)