

# How to Manage Risk in an Era of Explosive Unstructured Data Growth

Gain a deeper understanding of how to discover,  
classify, and continually manage unstructured data.



# Table of Contents

## ***PART 1 – Why data risk management is essential***

- Redefining data risk for today's era of heightened data protection
- What is the financial risk of data risk?

## ***PART 2 – Managing risk in unstructured data***

- What organizations need to successfully manage data risk
- Four key capabilities to properly manage unstructured data

## ***PART 3 – How unstructured data management works***

- Unstructured Data Discovery & Classification
- Unstructured Data Governance
- Intelligent File migration

## ***PART 4 – Unstructured data management benefits***

- The benefits of continuous unstructured data risk management
- Free scanning assessment

# Part 1 – Why data risk management is essential

## Redefining data risk for today's era of heightened data protection



Data risk is traditionally defined as the exposure to loss of value or reputation caused by issues or limitations to an organization's ability to acquire, store, transform, move, and use its data assets. With the rapid evolution of data protection regulations, like the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) in the US, the data risk definition needs to evolve accordingly. A more current definition of data risk is the exposure to significant financial damage due to data loss, business downtime, regulation non-compliance, or loss of reputation due to poor data security and/or data privacy practices.

The estimated amount of [data created daily is 1145 petabytes per day](#), which is expected to grow to 463,000 petabytes (or 463 exabytes) daily by 2025. Enterprises are becoming increasingly overwhelmed with the amount of data being produced. Understanding this data is difficult due to the extreme volume and its complexity. IT teams need automation to correctly identify, classify, and manage data at this scale.



### What is the financial risk of data risk?

Fundamentally, data risk is financial risk. Not having a comprehensive understanding of enterprise data brings a host of adverse outcomes that comprise data risk. You aren't fully compliant with data privacy regulations if you don't have complete control over your data. This also means that critical data may not be adequately secured, which brings the added risk of damage to the business from data theft, corruption, or loss.



**Here are the most common data risks and their financial consequences that can be caused by a lack of unstructured data management:**

Data Risk	Financial Risk
Having sensitive data that is improperly secured	If this data is stolen, corrupted, or encrypted it can cause a host of financial damage including: business downtime, regulatory fines, loss of reputation/revenue, and costs to repair breach.
Data that is exposed by employees, intentionally or unintentionally	A significant threat to data risk is not malicious actors, it's actually your employees. IT sets permissions rules, but employees find workarounds in order to get their jobs done. They don't intentionally mean to cause harm by sharing a payroll list to a third party, but the end result is a privacy violation which can cause reputation damage and regulatory fines.
Not knowing where personally identifiable information (PII) exists across systems	This will cost money in the form of resources needed to find data when it is requested to be pulled on an individual for regulations such as the California Consumer Privacy Act (CCPA). Additionally, this increases the risk of having to pay fines for non-compliance. Even unintentional violations could cost organizations \$2,500 per infraction.
Having too much stale, obsolete and unused data	There is an inherent belief that all data has value. This is why we stopped deleting emails and why organizations archive instead of deleting files. The risk here is twofold: First, with the explosion of data production, storage costs are increasing both on-premises and in the cloud. The second risk is the category of data breach and regulatory non-compliance. Just because you aren't using the data, doesn't mean it can't hurt your business if it is exposed—particularly if it contains PII.
Inability to provide a clear and comprehensive picture of data protection to auditors	The number of active data security and privacy regulations is growing rapidly and enterprises need to ensure auditors that they're taking steps to maintain compliance. Failure to do so can lead to more significant fines. Since it went into effect in 2018, the GDPR has assessed more than 1,600 fines, worth more more than \$4.3 billion cumulatively.

# Part 2 – Managing Risk in Unstructured Data

## What organizations need to successfully manage data risk



Traditional enterprise data management solutions have been around for decades. Most organizations have solutions that help them construct and maintain a framework for ingesting, storing, mining, and archiving the data integral to their business. Unfortunately, there is a crucial element missing from these legacy solutions: the ability to classify and manage their unstructured data.

Organizations need better methods of analyzing their unstructured data, identifying potential vulnerabilities, and automatically mitigating data risk. As data volumes continue to grow, there's a corresponding need for solutions to help organizations address the growing demand.



### Four key capabilities to manage unstructured data

#### 1. Discover and classify unstructured data in any system

Organizations have a ton of data, and not all of it is known, let alone visible for risk analysis. Dark data is typically referred to as all of the unused and unknown data that an enterprise generates. This accounts for everything from server log files to the multiple versions of documents employees create while doing their daily work. Unstructured data is defined as information that is not arranged according to a pre-set data model. Examples of unstructured data include texts, email, video, photos, webpages, audio files, and many business documents. According to IDC, 90 percent of the data businesses generate and collect is unstructured.

Due to the nature of unstructured data, it's incredibly challenging to identify whether the files contain any sensitive information, such as PII, or whether the information has been shared externally. Enterprises need better visibility into this mass of largely undiscovered data to properly protect their data.

The first key capability for proper data risk management is discovering unstructured data in any storage system, such as Microsoft OneDrive, Google Drive, Box, Dropbox, or on-premises file servers. It is critical that the unstructured data management solution can be fully integrated across your environment for this purpose. This helps eliminate disparate data management tools and ensures uniform data management across the organization.

The second key capability is Artificial Intelligence-driven (AI) unstructured data classification. A platform that employs AI is able to analyze the documents and accurately identify whether they contain sensitive information at a speed and scale that humans cannot match. An important feature to look for is pre-trained AI technology, meaning it already knows how to identify standard data and document types like invoices or resumes. This makes the solution magnitudes more powerful and provides a faster time to value than a solution that has to be trained in your environment before you can use it.

## **2. The ability to reduce the data storage footprint and costs**

As unstructured data volumes grow, so does the cost of storage—on-premises and in the cloud. According to Veritas, [over 33 percent of an organization's data](#) is estimated to be redundant, obsolete, or trivial. Storage cost is becoming a concerning line item for IT teams, and any capability that can allow them to reduce data risk while cutting costs is a big win.

Using an AI-driven unstructured data management solution to identify and manage stale data allows IT to treat that data as cold and archive it appropriately or to delete it altogether.



### **3. Automated controls that mitigate human error**

Data risk management treads a fine line between data protection and business productivity. Employees can't get their jobs done if you lock a system down too tightly with limited permissions. Too loose, and you're non-compliant. Another challenge is enforcing guidelines if users have complete control over data access. IT teams need an unstructured data management solution that helps them better control access to documents and files in an automated way, so IT can focus on keeping employees productive without worrying about regulatory compliance.

Beyond automating data access controls, organizations need a solution that can automatically migrate data to appropriate storage locations based on its classification. For instance, human resources files containing sensitive PII should be transferred to a highly secure location with limited access. Automating this action via a scheduled, rules-based policy is highly valuable to IT teams and can mitigate the risk of data getting in the wrong hands.

### **4. The ability to seamlessly handle reorganization events**

In large enterprises, reorganizations are common and require their own level of data risk management, but few things are more disruptive than a merger or acquisition. Acquiring organizations need to quickly scan and classify the data from the merged organization to prevent taking on unknown risks. They also need the ability to automate data migration into their environment, especially in a way that can assess and mitigate risk along the way.

In the case of a divestiture, it is critical to identify the enterprise content that stays with the business to prevent the loss of data and intellectual property. In organizations and industries with prevalent M&A activity, it is extremely valuable to employ holistic unstructured data management solutions that decrease risk and increase efficiency, with minimal end-user disruption.

# Part 3 – How it Works



## How unstructured data management works

Many organizations need help with the overwhelming complexity of unstructured data. According to IDC, as much as 90% of all data is unstructured and growing at unprecedented rates.

Traditional solutions aren't fast, flexible, or scalable enough to keep up with data growth or the ever-evolving regulatory environment. But modern approaches offer a better path forward.



# The three key functions of unstructured data management explained

## 1. Unstructured Data Discovery & Classification

To better understand an organization's data, modern unstructured data management platforms will automatically discover file properties such as:

- Type (like pdf or jpg)
- Size
- Age
- Location (storage system and file hierarchy)
- Internal and external access and sharing permissions
- Custom metadata fields
- Whether the file contains sensitive information

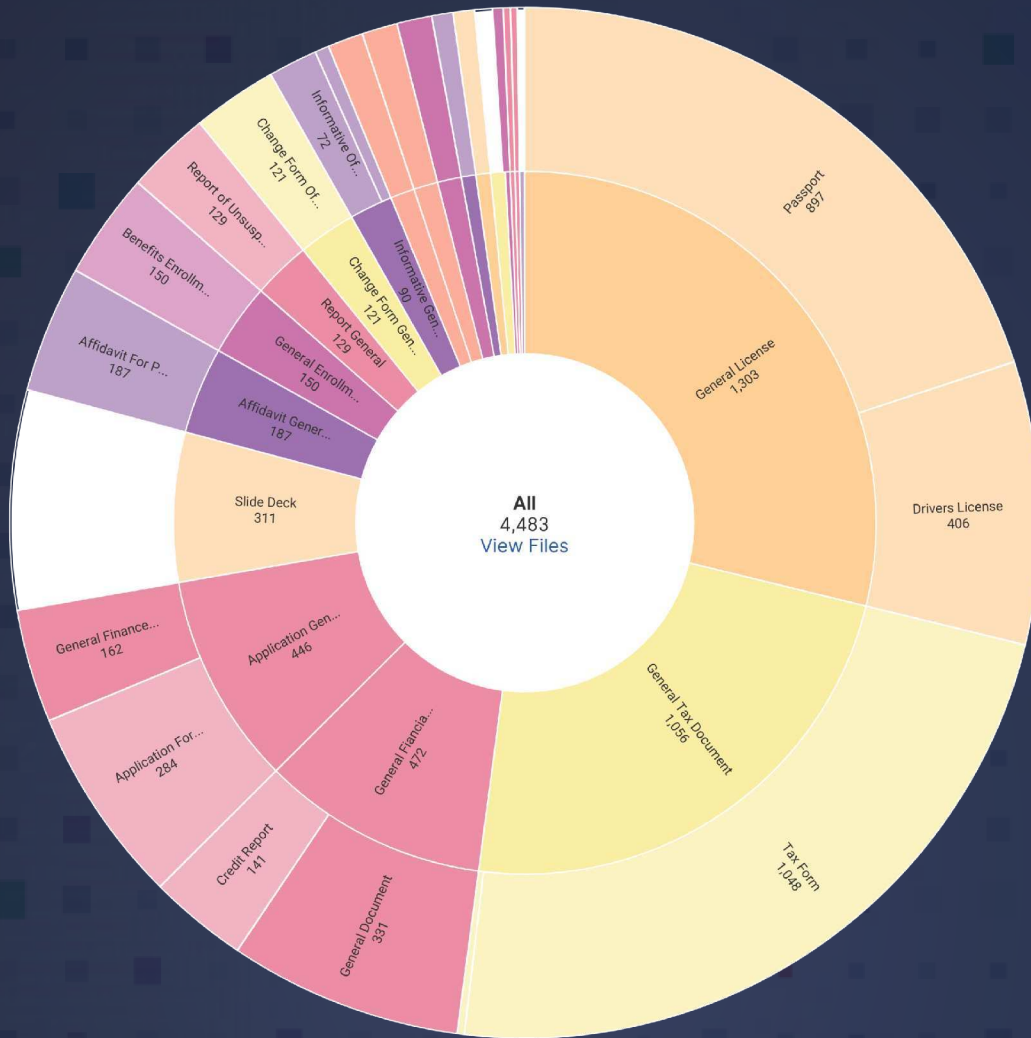
The platform must be integrated across all enterprise content repositories and storage systems to do this discovery. This provides an overview of what type of data you have, where it's stored, and who has access.

Once connected to the repository, the platform will automatically classify your content by scanning the documents and extracting the text and content for analysis to determine the document type and whether it contains sensitive information. Pre-trained AI is used to review and compare your unstructured data to known data types using advanced pattern matching. It automatically can identify what it is, assess its sensitivity, and apply a classification label.

AI-driven unstructured data management platforms should be able to classify the following:

- Document type (resume, W-2, invoice, etc.)
- PII, including names, ages, addresses, dates of birth, phone numbers, social security numbers, banking information, etc.
- Standard government forms
- Foreign language detection
- Any custom data attributes unique to an organization

## Example of data classification involved in data risk management:



## 2. Unstructured Data Management

Advanced unstructured data management solutions go beyond discovery and classification by automatically and continually performing actions to manage and protect enterprise data.

A series of actions can be taken based on the data discovery and classification results. These actions can be automated as organizations continually generate and update their data. Through activities like transferring data to new locations, re-applying classification labels, or other modifications based on sensitivity and risk, continual data management enables an always-on risk management approach.



Actions that can be triggered to automatically manage and protect data:

- Flag or re-classify the data
- Transfer to a new storage location
- Apply universal persistent labeling
- Apply Microsoft MIP (Purview) labeling
- Quarantine the data for a set period of time
- Change permissions, including sharing and access
- Archive or delete data

Actions can be automated at a file level, but the information around data risk across the organization can also be aggregated to make larger decisions to enhance data security, increase operational efficiencies, and reduce the storage footprint.

The most exciting innovation in modern data management is leveraging artificial intelligence processing to continuously monitor enterprise content for sensitive data or incorrectly applied user-defined classification labels that may expose the business to financial risk from data loss or non-compliance.

The unstructured data management platform typically presents results via a risk assessment dashboard or in custom reports that reveal any sensitive data found, and any automated actions taken to mitigate that risk.

Unstructured data management provides continuous oversight to:

- Ensure data is classified for both protection and discoverability
- Ensure data is labeled so that DLP policies can be applied
- Ensure that permissions are set correctly so that sensitive data is not exposed to unauthorized users



**Nicolet**  
NATIONAL BANK

*See how DryvIQ provided an immediate remediation plan for Nicolet National Bank that has automated ongoing policy enforcement and continually discovers and protects data as it's updated or created.*

[Read More](#)



### 3. Intelligent file migration

Data migration can be a part of regular data hygiene, whether in the form of a large-scale migration to a new repository or an AI-driven approach that selectively migrates content based on document classification, age, or type. Modern unstructured data management platforms have both of these capabilities.

Intelligent migration capitalizes on the unstructured data management platform's ability to surface greater insight into your data to execute a more efficient and lower-risk migration.

Before moving any data, intelligent file migration helps answer questions like:

- How much content are we managing? How old is it?
- What types of documents do we have and are they sensitive?
- Who can access my content? Has it been shared externally?
- Have end users have been properly labeling content manually?
- Do we actually need to migrate all this data?



# Part 4 – Unstructured Data Management Benefits



## The benefits of continuous unstructured data management

Today's organizations face rapidly growing data volumes, changing regulations, and limited resources. These challenges are driving data initiatives across the enterprise. By taking a modern approach to managing unstructured data, organizations will be better equipped to keep up with these demands, with greater visibility into how they can protect their data and extract more value from it.

<b>Lower risk of data loss</b>	<b>Increased Efficiency</b>	<b>Reduce operational costs</b>
Being able to identify sensitive data and secure it appropriately mitigates the threat of data loss. It also brings investor confidence in the org's cyber resiliency.	Automating data classification and policy enforcement eliminates manual effort and enables orgs to keep up with data growth rates while improving productivity.	Better data management can reduce storage fees and free up IT resources to optimize systems and processes to cut costs and lower risk.

# Scan & classify your unstructured data

The DryvIQ platform offers a single solution to continually classify, manage, and migrate unstructured data across more than 40 on-premises and cloud-based platforms — at speeds of up to 100TB per day, at a petabyte scale. More than 1,100 worldwide organizations have put their trust in DryvIQ to help them address their large-scale, complex unstructured data management projects.

If you're ready to take control of your unstructured data but aren't sure where to begin, **contact us for a free scanning assessment.**

[Get Your Free Assessment](#)