CYERA

EBOOK

How to Overcome Common DLP Challenges with Cyera's AI-Powered Data Security Platform

The benefits of cloud-native, data-centric security to harden your security posture and assure compliance across IaaS, PaaS and SaaS

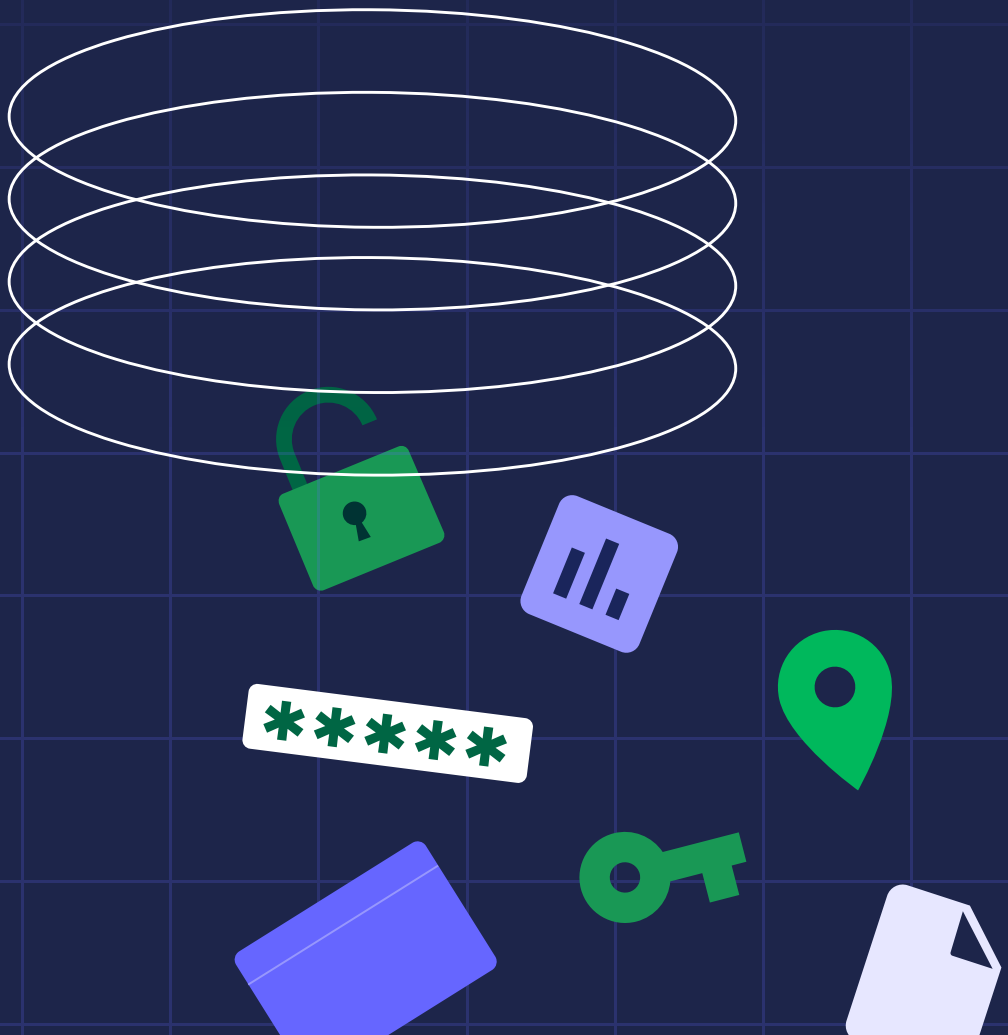


Table Of Contents

Summary 03

Why Traditional DLP Tools Lack An Accurate Data Foundation 04

Why It's Getting Even Harder to Trust Traditional DLP Solutions 07

- The Shift to Cloud Infrastructure 07
- A Lack of Alternatives 07
- Limited Innovation in the DLP Industry 08

Choosing A Data Security Platform to Complement Your DLP tool 08

How Cyera Enables Organizations to Operationalize DLP 09

- Operationalizing Your DLP Tool Using Cyera 09
- Improving Your Data Security Posture with Cyera 10

About Cyera 10

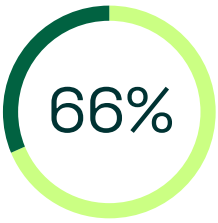
How to Overcome Common DLP Challenges with Cyera's AI-Powered Data Security Platform

Implementing a data loss prevention (DLP) solution is challenging for many organizations, and even the most mature security programs struggle to realize the full capabilities of these tools. Moving from basic monitoring to true data loss prevention — where data activities are allowed or blocked automatically by the DLP system — requires a high level of confidence in the accuracy of their data discovery and classification processes.

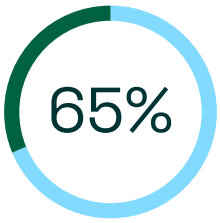
The problem is security teams don't trust that the manually configured, point-in-time capabilities built into traditional DLPs provide an accurate data foundation to begin blocking activities without false positives, or detections devoid of context disrupting the business.

In this eBook, we'll discuss the problems with traditional DLP tools, why it's getting even harder to trust DLP solutions in today's cloud-first data landscape, and what organizations can do to improve their data security posture.

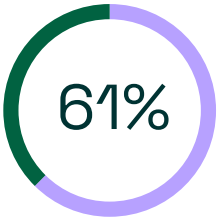
> **Information Security leaders indicate that improving data risk assessment accuracy (82%), data security posture in the cloud (79%), and increasing automation of security controls (76%) are critical or high priorities.**



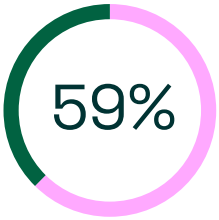
Struggle to identify security exposures for sensitive data



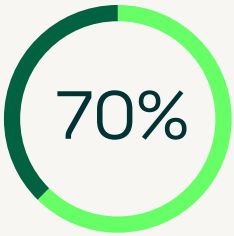
Struggle to enable controls to protect data



Struggle to identify compliance violations



Struggle to maintain detailed data inventory



Struggle to enable the business to use the data

Why Traditional DLP Tools Lack An Accurate Data Foundation

Many organizations adopt an enterprise DLP to protect their sensitive information, and quickly discover that they can't use the "loss prevention" capabilities without impacting business activities or wasting IT resources. Despite many of these tools specializing in specific aspects of DLP — such as email, endpoint, cloud, or network — they aren't designed to intelligently understand your sensitive data landscape.

Without a solid data foundation, it's nearly impossible to write effective DLP policies. This means traditional DLP tools generate too many false positive alerts due to policies being tied to an incomplete or inaccurate understanding of the data. Besides false positives, ineffective DLP policies can also lead to inadequate protection and other data security challenges.

Here are four limitations that prevent DLPs from achieving their primary goal of stopping sensitive data from leaving secure environments.

1. Manual Data Discovery

Summary

- **47% of information security leaders say manual data security processes are cumbersome and time-consuming**
- **Manual data discovery means humans connect technology to the data they know about, providing a limited view of data**
- **Environment and data store support is limited, requiring multiple overlapping tools that create silos of understanding**

The built-in data discovery functionality of DLP tools, at best, identifies the datastores you already know about. That's because DLP tools either require implementers to manually connect them to the datastores they are intended to protect, or can only protect a very limited subset of the datastores a business manages. In fact, 47% of information security leaders find that existing approaches to data security are overly cumbersome and time-consuming.

Most DLP data discovery features do not automatically uncover unknown datastores or detect changes in existing ones. They also are limited in the environments or deployment models they support, providing siloed views of data that are incapable of ensuring real data loss prevention. This means security teams need to implement multiple tools and manually trigger scans to capture changes in their data environments, often preventing organizations from keeping up with today's rapid pace of data generation.

Amazon Macie for data discovery

AWS's native data discovery tool provides limited support for AWS data stores

Supported	Not Supported	
S3 Buckets	DocumentDB	AWS RDS
	Redshift	AWS EC2
	DynamoDB	

Figure 1 - Example of limited datastore support with cloud DLP tool Amazon Macie. Tools by Azure and Google Cloud have similar limitations.

2. Rule-Based Data Classification

Summary

- 42% of information security leaders say implementing data security technologies takes too long
- Rule-based data classification uses complex regular expressions to detect static patterns in strings, requiring constant maintenance and tuning
- Regular expressions result in a rudimentary understanding of data, devoid of context, that result in inaccurate, noisy alerts

Traditional DLP tools have a limited understanding of where data is managed, and its level of sensitivity because they use regular expression-based (RegEx) classification. RegExes are a concise and flexible tool for describing patterns in strings. They are useful because strings usually contain unstructured or semi-structured data, but it is important to recognize that they are simply a tool to find static patterns.

RegExes are notoriously difficult to implement. Not only are they hard to read, but relevant examples are hard to search for, hard to validate, and hard to document. Writing effective RegExes requires special skills to avoid false positive (incorrectly recognizing a pattern in a string) and false negative (missing a valid pattern in a string) classifications. RegExes also tend to be very dense. The format of complex regular expressions has been compared to what it might look like if a cat walked across the keyboard.

For these reasons, troubleshooting RegEx classification is one of the most time-consuming aspects of implementing DLP solutions. They also result in only a rudimentary understanding of what data represents, devoid of any context. When metadata, cloud tags, and sensitivity labels are implemented based on RegEx classification, they result in large volumes of inaccurate, noisy alerts. This is the most common reason that DLP solutions are implemented in alert-only mode - security teams cannot risk disrupting the business.

3. Lack of Contextual Awareness

Summary

- 65% of information security leaders struggle to enable controls to protect data
- DLP solutions claim to develop data context, but rely on inconsistent, static metadata and cloud tags, typically defined by static RegEx classification
- Without information on the environment, identifiability, security, and access to data, DLP policies are either ineffective or disrupt the business

DLP tools claim to provide context on the data they classify. However this context is derived from metadata and cloud tags, which the previous section identified are fraught with challenges. Common sources of data context — such as cloud sensitivity tags and Microsoft Information Protection (MIP) sensitivity labels — are inconsistently applied, only represent a point in time, and lack proper awareness, education, and governance as environments, applications, and teams evolve. DLP tools that rely on these legacy data classification methods usually require human oversight to manually review and validate data labels on a regular basis.

As a result, DLP policies end up treating data the same way without considering additional information about the data. This leads to ineffective policies and controls. According to Forrester, 65% of information security leaders struggle to enable controls to protect data. This is due to a lack of context, which results in a number of data security challenges:

- False positives or data leak alerts that are not actual security violations.
- False negatives where toxic combinations of data aren't recognized.
- Overly restrictive data sharing that blocks actions that should be allowed.

The reality is, the same data will typically benefit from a variety of DLP policies depending on the characteristics that define the data, information about the environment where it's stored, the controls in place to ensure data security and integrity, and the framework that regulates the data. When DLPs aren't applied based on these real-world circumstances, this can lead to cybersecurity risks, compliance violations, and a loss of trust.

Not having context also means that administrators who write policies don't have enough information to decide who should have access to the data. In turn, the DLP might be overly restrictive and prevent users from doing their job, or not restrictive enough and allow sensitive information to leave the organization. Without context about why users are accessing data, many DLP policies either disrupt business processes or are ineffective at preventing data loss.

Why It's Getting Even Harder to Trust Traditional DLP Solutions

> **Forrester found that for 98% of information security leaders, the data security status quo at their company is a problem.**

Although the lack of an accurate data foundation has been an issue for a long time, it's getting even harder to trust traditional DLP solutions each year. The shift to cloud infrastructure, a lack of alternative solutions, and very little innovation in the industry have limited the effectiveness of traditional DLP tools.

The Shift to Cloud Infrastructure

In modern cloud environments without traditional network borders, it has become more difficult to protect data at rest, in motion, and in use. While traditional DLP solutions were effective because emails, text documents, images, and other files were within a well-defined perimeter, enterprise infrastructure is a complex web of cloud services, SaaS applications, APIs, and more. Since there is no longer a network perimeter, and modern microservices architectures are designed to cross cloud provider boundaries, it's no longer relevant or possible to protect data that's "leaving" an environment.

The cloud has also caused companies to generate large volumes of raw data. Yet traditional DLP tools can't easily protect massive amounts of unregulated and unstructured data because they don't have enough contextual understanding to correctly apply security policies. Moreover, the siloed nature of DLP tools mean they cannot effectively identify when data moves across environments, such as between cloud Infrastructure and SaaS applications. Organizations need a data security solution to continuously monitor for changes to keep up with the dynamic nature of how data is generated, consumed, and used.

A Lack of Alternatives

Many organizations try alternative approaches to operationalize their DLP tools, but they are usually inadequate. Here are two common techniques:

- **Exact data matching (EDM)** takes a large database of all known information in your environment and matches it to outbound data transfers. This gives you a higher level of confidence when blocking or allowing data activities, but you still need an accurate data foundation.
- **Indexing involves pointing your DLP system** to a data store and creating a fuzzy hash of those files to be able to identify them with a higher level of confidence. However, you still need to identify all of your data stores to implement indexing.

Techniques like exact data matching and indexing might slightly improve the effectiveness of a traditional DLP tool, but organizations still need cloud-native data security solutions that can truly prevent data loss.

Limited Innovation in the DLP Industry

All of the challenges mentioned previously continue to persist because there's a lack of innovation in the DLP industry. Organizations that rely solely on legacy DLP solutions find themselves with incomplete protection against data loss in today's cloud-first data landscape. This introduces security and compliance risks that could easily result in data breaches, regulatory penalties, and reputational damages.

While some DLP companies are slowly innovating and adopting new technologies like AI and machine learning, the fundamental architecture is built on top of manual processes that require lengthy implementation cycles and continuous maintenance. This is why Forrester found that 71% of information security leaders expect transformational benefits from automating data security.

Organizations that want to operationalize their existing DLP tools will need to look beyond the traditional DLP industry for modern security solutions that are purpose built for the cloud era, and capable of keeping pace with advances in technology like generative AI. This requires a high level of automation, rapid time to value, and the layering of multiple technologies, including pattern matching, machine learning (ML), and Large Language Models (LLMs).

> **Security leaders are investing in modern data security technology to drive meaningful change including dynamic security controls (81%), real-time exposure detection (76%), and data security posture management (72%).**

Choosing A Data Security Platform to Complement Your DLP tool

As you can see, relying solely on a traditional DLP solution isn't enough in today's cloud-first data landscape. Here are some key features to look for when choosing a data security platform to complement your existing DLP tool.

1. Instant visibility of data

An effective data security solution shouldn't take weeks or months to provide an understanding of what sensitive data you have and where it lives. Implementing a DLP requires in-depth visibility into your data so that you can write DLP policies to start protecting data. This means the process for building an inventory of sensitive data should be as frictionless as possible. To achieve this, start with a data security platform with agentless connections to your data sources.

2. Sensitive data discovery

Most traditional DLP solutions require administrators to point the data discovery system to any known datastores. With organizations generating more and more data each year, it's no longer feasible to manually identify every data store. There needs to be automated data discovery capabilities that can uncover unknown data stores for DLP tools to adequately protect all data

3. AI-powered classification

Since most traditional DLPs rely on rule-based classification systems, they require a lot of tuning just to correctly label a subset of data. With the proliferation of data and different categories of data, however, it is essential to adopt an AI-powered classification that can accurately classify data, at scale, with minimal to no tuning.

How Cyera Enables Organizations to Operationalize DLP

Cyera is a holistic data security platform that can help protect your data in the modern, cloud era. By automatically building a modern, purpose-built data foundation with Cyera, your security team will be able to write more effective, dynamic DLP policies that reduce alert fatigue and increase data security.

More specifically, Cyera implements data discovery and classification better than exact data matching, indexing, and other approaches to operationalize legacy DLP tools in a cloud-native environment. This means Cyera provides a high level of trust to fully utilize the prevention capabilities of traditional DLP solutions without generating a lot of false positives, false negatives, or overly restrictive data-sharing policies.

Operationalizing Your DLP Tool Using Cyera

Here's how to operationalize your traditional DLP solution using Cyera:

1. Automatically generate a view of sensitive data

Cyera automatically discovers your data sources across different datastores, including cloud storage buckets, databases, containers, virtual machines, SaaS applications, and more. This ensures you have a complete view of your data, even in fragmented and sprawling cloud environments.

Then Cyera classifies your data based on classes, cloud tags, and MIP sensitivity labels with a deep contextual understanding. Cyera also uses advanced AI/ML and large language models for advanced data classification. This creates a holistic picture of your data, laying the foundation for more optimal DLP policies.

2. Write DLP policies that address the intended data

Using Cyera's accurate data foundation, your security team can write policies that fit your specific data requirements in a way that reduces false positives. You'll be able to create DLP policies that consider where the data is stored, the sensitivity of the data, the context in which it's being accessed, and other relevant factors.

Cyera can also get your Microsoft instance to automatically write MIP tags, which almost every legacy DLP tool can read. This streamlines the process of writing DLP policies and integrating Cyera's data foundation with leading DLP products like Zscaler, Netskope, and Forcepoint. In turn, you'll have security policies that better align with your data, giving your security team confidence to automatically block certain data activities.

3. Optimize DLP policies by monitoring changes to data

Although your data is constantly changing, Cyera will continue to scan your data stores and adapt to reflect its current state. This helps overcome the limitations of DLP tools with legacy data discovery and classification processes, which require manual effort and are slow to adapt to changes.

DLP policies tuned to current data minimize false positives and alert fatigue in the long run. If data becomes more sensitive over time, then you can implement more restrictive DLP policies. You'll also be able to protect new data stores as soon as they're created to stay ahead of potential security risks.

Improving Your Data Security Posture with Cyera

Besides enabling traditional DLP tools to better protect data, Cyera has additional data security capabilities that improve cyber-resilience and compliance. For example, data security posture management (DSPM) capabilities allow organizations to prioritize remediation efforts to help security teams elevate their overall security posture.

In addition, Cyera provides automated remediate workflows based on established security risk, regulatory, and compliance frameworks. By simplifying incident response, triaging, and compliance audits, Cyera also increases the productivity of your security team. Combining powerful data discovery and classification, data access governance, and other data security capabilities together in a single platform empowers organizations to leverage their data securely.

Trust is at the core of operationalizing a traditional DLP, and it starts with a solid data foundation from Cyera. Security teams can write DLP policies to allow and block certain data activities with a high level of confidence when they have a deep contextual understanding of their data landscape. This is the key to achieving a lasting improvement in data security.

Want to learn more about operationalizing your traditional DLP solution with Cyera?

Contact us or schedule a demo today

Get a Demo

Cyera commissioned Forrester Consulting to execute a study that surveyed 353 information security decision-makers at North American-based companies about their data security priorities, challenges, and technology investment plans. The research was conducted in August 2023, and the full report is available via cyera.com/resources.

About Cyera

Cyera is the data security company that gives businesses deep context on their data, applying proper, continuous controls to assure cyber-resilience and compliance. Cyera takes a data-centric approach to security across your data landscape, empowering security teams to know where their data is, what exposes it to risk, and take immediate action to remediate exposures. Backed by leading investors, including Sequoia, Accel, Cyberstarts and Redpoint Ventures, Cyera is redefining how companies secure data in the cloud.

To learn more, visit www.cyera.io

Trusted By

LifeLabs ACV Cboe Mercury Financial Veolia Cystic Fibrosis Foundation ADVANCE UTA