FORTRA

WHITE PAPER *(Titus)*

5 Reasons Classification is the First Step to Successful Data Loss Prevention

Introduction

Organizations of all sizes and across all industries are facing the ever increasing threat of data breaches. Whether by malicious attack or accident, exposed confidential information can place an organization at risk of fines, lawsuits, lost revenue, and damage to its reputation. To prevent a data breach from occurring, organizations are turning to Data Loss Prevention (DLP) solutions, and more recently their Cloud counterparts, Cloud Access Security Brokers (CASB). Yet, many organizations that have adopted DLP remain at risk.

While incredibly powerful, DLP solutions have difficulty in correctly identifying the various types of data they scan. To determine data's sensitivity to the organization, DLP systems rely on search algorithms which have poor multimedia scanning capabilities and lack the means to determine the context of the data. Without explicit classification to signal proper governance procedures, DLP systems either catch too much (slowing the flow of business and frustrating workers) or too little (allowing sensitive data to escape). As a result, many organizations running DLP turn off the "prevention" tools and simply monitor data traffic, effectively reducing them to the same, pre-DLP status of reacting after a breach has occurred, albeit with much better auditing capabilities.

As the importance of protecting data has grown, various technologies have improved their native ability to prevent data loss. Use of the native data loss prevention capabilities within security and productivity applications is providing an alternative to the implementation of a dedicated DLP suite. However, the multiple solutions that make up a collective DLP strategy face the same challenge – they have to know how sensitive the data is in order to apply the appropriate policies and protections.

In this white paper we will review five reasons why data classification is an essential first step to achieving the maximum return on investment for any DLP implementation. By identifying and classifying data, DLP systems can run more efficiently, accurately protecting data while freeing system administrators from excessive manual review.

Data Loss Prevention Missteps

Although the digital world has made it easier to work, share, create, and collaborate, it has also made it much easier to gain access to confidential and private information. We hear data breach stories daily, be they "whistleblower" leaks, theft of credit card information, accidental emails containing medical records, or lawsuits over copyright infringement. Public and private organizations across the globe must constantly be on guard against internal and external threats as data exposure negatively impacts reputation, revenue, public safety, financial stability, and market share. Data security has therefore become one of the top priorities of any organization, regardless of size.

Rising to meet the requirement to provide better digital data governance, an entire industry has evolved, and there are now many products and technologies dedicated to data protection. There is no single product, however, that provides a complete solution to the complex data security problem. There are solutions that provide encryption, others that manage collaboration, and a niche dedicated to rights management. Suffice it to say, "some assembly is required" whenbuilding a complete security structure. Yet, without a clear indication of what technology should be the security cornerstone, organizations must find their own way.

FINDING THE STARTING LINE

Determining the starting point for a data loss prevention project is influenced by several factors, including:

- budget
- infrastructure and IT resources
- business and regulatory requirements
- corporate security priorities
- user acceptance and ease of use

Usually, the choice most organizations arrive at—influenced by a great deal of market hype—is that implementing a DLP system is the correct first step. And on the surface, the reasons for this choice are indeed compelling. DLP solutions are exceptionally powerful security tools; they can monitor, detect and block sensitive information from traveling over the network, through endpoints, and protect archived data, all according to corporate policies. DLP systems are the heart of any data security initiative but, as powerful as they are, they lack the means to provide the needed identification accuracy on the data they process. This can result in blocked non-sensitive data (false positives) or mistakenly released sensitive data (false negatives).

In most cases, IT departments become overwhelmed with the volume of reports and alerts DLP systems generate. Likewise, business process owners become frustrated when "Big Brother" is constantly blocking information flow, hindering their ability to conduct business. With such high human resource and business costs, many implementations of DLP are nearly shut down. User and administrator frustration inevitably leads to the weakening of DLP data blocking algorithms, or to restricting data blocking to one type of information, or—in some cases—no blocking at all. According to Forrester research, about half of all companies feel that their DLP implementations have failed at some level, despite the high expectations and sizable investment.[1]

Some DLP implementations are reduced to simple traffic monitoring so that, when a security breach does occur, it can be accurately audited. Unless your DLP system is actively and accurately blocking the release of confidential information, your investment in DLP is being squandered.

> ***"Most security teams fail to achieve DLP success because they don't define the necessary process and policies before their deployment. DLP tools are not 'automagical.' They can't find data if they don't know what to look for. Security professionals must train DLP tools by defining policies, but before you can define policies, you have to properly inventory and classify your sensitive information."***
>
> **— John Kindervag**
> Principal Analyst, Forrester Research

THE FIRST STEP

When looking at best practice documentation discussing successful data protection, the first phase is usually discovery and identification. It makes sense; you need to know what data you have to properly protect it and create the rules needed to govern its distribution. It is worth noting, however, that "discovery" does not necessarily mean that a complete inventory of your unstructured data needs to occur before a DLP program can be considered a success. It is recommended that data identification and classification initiatives begin with new data.

Why?

Because new data contains the most valuable information. Take, for example, information such as the latest innovation from the R&D department, this quarter's sales figures, and current merger and acquisition analysis. The value of this information will drop considerably as it ages. Secret product plans, financial performance data, and merger information is much less valuable once the innovative new product is in stores, the annual report is published, and the merger complete.

In addition, new data is more likely to be accidentally breached as it is in constant motion, being shared between multiple users across networks, email, mobile devices, and cloud repositories. In contrast, the value of old data is diminished as it is no longer part of the business workflow and typically sits at rest, rarely accessed, waiting to be deleted.

Another reason to begin with a dedicated classification solution is because many DLP solutions have no way to permanently identify the data. If, for example, the data is accessed, moved, or copied, the DLP system must run the detection algorithms a second time to determine the data's sensitivity to manage it according to policy. Furthermore, starting with a discovery project often delays progress due to the sheer volume of data. For many organizations, reviewing the huge stores of legacy data is a daunting and slow process as retroactively identifying your data collections involves automated searches that deal only in probabilities, not in absolutes. Identification discovery searches are only as good as their search algorithms. As a result, the search may not correctly categorize the data without considerable time and effort placed into making sure the identification and subsequent classifications are accurate.

If the data of greatest value—and therefore at greatest risk—is the data that is being used, why hinder your project by starting with the data that is least at risk and the most difficult to identify? Instead, consider focusing on active data which can be categorized or classified by the users while they are working with it. Data and document classification tools provide users the means to apply security, handling, and otherwise identifying labels to documents and files, enabling the DLP to work with absolutes. Data classification provides permanent and explicit identification labels DLP systems need to process the data correctly.

Let's take a closer look at why organizations who want to get the most from their DLP implementation choose to roll out classification first.

Top 5 Reasons to Classify First

REASON 1 | DATA SECURITY IS A BUSINESS PROBLEM THAT TECHNOLOGY ALONE CANNOT SOLVE

There is a widely held belief (or perhaps simply a hope?) that data security can be solved by implementing a new piece of technology. Stopping data from being downloaded, encrypting data, ensuring access credentials—all of these protections can be programed into a security net designed to prevent breaches. True security, however, is a constant process that involves everyone in an organization. Exclusive reliance on automated systems will doom your project to failure.

Many DLP implementations hit their first snag with the initial setup. Often, the IT department is given a list of criteria that defines sensitive information and security policies for dealing with it. Beyond defining what the DLP system must look for, the data and business process owners are not involved in enforcement. Even though users are a large part of the problem (either through accidental or malicious intent), they are not required to identify the data they are handling. The task of protection is left in the hands of IT administrators.

Handed their instructions, IT staff program the search algorithms that catch data breaches. Assuming they have accurately interpreted the instruction from the business process owners (that are often simply lobbed over the wall), IT creates rules for detecting and then managing data leaks. To ensure that nothing is leaked, these algorithms are set to be stringent at first, meaning that many potential breaches are caught. But, many "catches" are not security breaches at all. The tighter the security, the more "false positives" are caught and the more calls workers place to the IT department asking for data to be released. False positives are a big problem as they:

1. Require manual handling (review or release) by the IT team, and
2. Stop business workflow, frustrating the users

The IT department is ill suited for the task of determining what constitutes a breach and what does not. It overloads them with added work and, in some cases, this review by IT may itself constitute a security breach. Without user involvement, DLP systems are guessing on the sensitivity of the data. If the users had means to tell the DLP system how to handle the data, IT would not be put in the position of having to review excessive data breach reports or have to respond to constant requests from information owners to let their data go.

Business user frustration is another negative side effect of making data security an IT issue.

> ***"Organizations are continually sharing data with partners, clients, and even competitors in this age of collaboration and transparency. There are significant benefits to involving the end users directly in the protection of data, and in turn in their organization's overall DLP strategy."***
>
> **— Eric Ouellet,**
> VP of Research, Gartner

Workers want and need to have the power to perform the tasks they were trained for and were hired to do. While it is important to prepare for the small fraction of individuals who may steal data, it is important not to treat your entire workforce as though you distrust them all. If the day your DLP system is turned on your workers find that activities they used to do as part of usual business practice are blocked or significantly hindered, there could be tremendous resistance and push-back. Even when they know the changes are coming, if the DLP system is catching too many false positives, the whole project could be at risk as angry employees harass IT to release their data or search for ways to circumvent security. The result? DLP security measures are weakened. Companies would rather deal with minimal data loss just to keep workers happy and the business rolling.

Users should be empowered to take responsibility for the security of data they use and create. User-driven classification provides much more accurate data identity and will thus help ensure the DLP system handles the data correctly. Greater accuracy will also release the IT team from excessive manual monitoring. User classification also has the added benefit of fostering a culture of security in the user community. Rather than being subject to "Big Brother," users are a respected part of the security solution that is in place to help protect their company and, subsequently, their jobs.

REASON 2 | CLASSIFICATION FOSTERS A SECURITY CULTURE

Security systems have done an excellent job at preventing prying eyes from gaining access to sensitive information in the corporate network. What they aren't as good at is preventing accidental disclosure by careless users with legitimate access. While a DLP's failure to catch a particular breach can be classified as an "error," it is the user who accessed and distributed the information that is the real problem. The act of asking (or forcing) users to classify each file while guiding them to correct decisions based on approved policy helps to improve the source of the problem: users who lack awareness of the proper security procedures.

Common data breach accidents include:

- Incorrectly addressed email
- Sensitive data in an email or email attachment
- Accessing data from unsecure, public sources
- Lost devices and storage media
- Accidental inclusion in e-discovery packages
- Inappropriate sharing to personal email and devices.

These breaches are predominantly caused by user ignorance or error. While a DLP system is vital to providing a second look when these mistakes occur, without classification not all breaches may be caught.

Despite all the time, money, and effort your organization may (or may not!) put into training staff on security policy and the proper handling of sensitive information, employees are not likely to retain this information to the degree necessary because they are not usually motivated by security. As work pressures ebb and flow, users tend to put security concerns aside to expedite business. Deadlines, commissions, being seen as efficient and as a hard worker; these are the motivations that drive most employees. They quickly forget why they need to protect information ("it won't hurt the company's profits") or they intentionally try to bypass security ("if I can't email this document I will just print it and take it with me") in their rush to finish a task.

[1] Kindervag, John. Rethinking DLP. Forrester Research, Inc., 2012. PDF.

Even if a DLP system does catch the breach, there is usually no informative response to help the user remediate or learn from their error. Depending on how the DLP system is configured, an email that violates the organization's security policies may be:

- Returned immediately to the user
- Put in quarantine pending manual review
- Encrypted and sent anyway (hopefully the recipient can decrypt)
- Automatically deleted

The user responsible for the email may not know for hours, days or ever that the email was blocked. Even if the email is sent back to the sender, the policy breach notification (normally just the policy rule name) may not contain enough details for the sender to know how to fix the email or avoid the same problem in the future.

This "solution" not only fails to prevent users from repeating the same error, but it creates frustration among the user community. Although the DLP system is there to help protect the users and the data they share, it becomes viewed as an impediment to business. In many cases, user push-back has even forced administrators to turn off data protection polices and simply rely on data monitoring. In monitoring mode, harmful data are not blocked; it is only recorded in logs. Pointing fingers after a data breach does nothing to mitigate the damage a breach can cause.

A classification tool, however, consistently reminds users of data security policies each time they save a document or send an email. By reminding (or forcing) users to identify the sensitivity of the information, data security remains constantly top of mind. TITUS classification solutions provide policy information to the user, guiding them through their decisions so they apply the proper classification designation. And, by checking the selected classification against the email content and attachments, classification tools can immediately identify possible breaches before the email ever leaves the user's control.

With classification, the user works with DLP and other security systems to ensure data protection policies are followed and enforced.

REASON 3 | DLP SYSTEMS HAVE TO KNOW THE DATA TO KNOW HOW TO MANAGE IT

To prevent data loss, your DLP technology must know what to block. DLP systems use powerful search algorithms to examine the data residing in, traveling through, and leaving your network. Based on what it finds, DLP systems have several options—from preventing access, to denying copy actions, to encrypting data. But all these useful data governance actions are dependent on how the DLP system identifies the data. Failure of the search algorithm means either failure to enforce the proper security policy or freezing the data until it is manually reviewed.

DLP searches look for key strings of text in the data or in its properties. In some cases, this data can be very specific, such as a Social Security Number (SSN). In other cases, the sensitive data indicators might be a specific string of text unique to your organization. In both cases, the DLP system is still making a guess; configuration of the DLP search algorithms determines how much is caught.

Some PII, such as credit card numbers, do have a precise mathematical formula which can be used by DLP systems for detection. But there are other items, like a SSN, where no validation algorithm exists, and as a result DLP search algorithms must be set to be fairly broad.

Take the following example of the nine-digit Social Security Number (SSN): 000-00-0000

A DLP system could be set to capture the specific sequence of numbers and symbols. The DLP could be specifically set to search for a group of three numbers, a group of two, and then a group of four numbers, all separated by a hyphen. Using this specific search criteria, the DLP would identify only data which contains this exact number/character sequence and would miss any others Social Security Numbers where the exact sequence was broken. If there were no hyphens, for example, the scan results would miss the number, resulting in a false negative. If the search rule is less stringent to include any sequence of nine numbers (not broken by more than one character or space between each number), then any and all nine number sequences—like telephone numbers—would be identified as

a security breach. The broader the search rule, however, the more false positives are found resulting in increased user frustration and review work.

Since false positives are such a burden, many organizations running a DLP system resort to loosening the reins. For instance, a policy may state that any email or document cannot be sent outside the company if it contains a Social Security Number. However, because of the number of false positives, the policy may be amended to stop only emails or documents that contain five or more Social Security Numbers. While this ensures the likelihood that the data does contain an SSN, it also means that small breaches are permitted.

Regardless of the content or the formatting, explicit classification metadata allows DLP systems to manage data with certainty. It doesn't matter if the DLP scan confuses a telephone number with a Social Security Number. Classification provides precise governance instructions in either case. Of note, the DLP system should still be configured to record when its scan conflicts with the classification. By using both tools, any irregularities in worker behavior can be tracked to locate careless or possibly malevolent employees.

Context is another area where DLP search algorithms cannot be relied on to correctly filter data. What might be sensitive information in one context may be innocuous in another. For example, sales data might be a closely guarded secret for a publicly traded company until the official earnings report is shared. In a different case, access rights to the information may have changed based on the user's role or even their physical location. Accounting for these context changes can be difficult to enforce programmatically. Yet, users know this information and should be given the ability to communicate context with the DLP gateway. In most cases, however, DLP administrators are left to decide the fate of quarantined data without knowing who it belongs to, its exact importance, or the intent of the sender. Once again, classification can solve this issue. Clear classification labels provided by a user who understands the current data context can provide unambiguous instructions the DLP can interpret for proper policy enforcement.

REASON 4 | DLP WORKS BEST ON KNOWN THREATS

DLP systems are designed to check for specific patterns in text. But, if the identifying data is difficult to isolate as risky (common phrases, shared terms) or is not text-based, DLP systems can miss this information all together.

Intellectual property (IP) often falls into the category of data that is difficult to recognize. Unlike a credit card number or a patient I.D., intellectual property is widely varying in format and is constantly being created faster than search terms can be updated. For instance, for each new project it may be required that DLP administrators create and test new rules based on the expected content. Without the new rules, the DLP system may fail to protect data about the new project.

Also, IP could take almost any form or media format. Chemical formulas, manufacturing processes, customer lists, product development documents; these are all examples of data that could either contain such specific terms that a DLP cannot realistically be updated to detect, or are so common that filtering to find them would bring up far too many false positives. Media files—such as videos, audio recordings and images—may contain private data or IP as well, but scanning their contents is difficult. Unless multimedia files are given an explicit classification using metadata the DLP can read, the DLP search capabilities are nearly powerless.

Potentially the most valuable asset to an organization, intellectual property must be protected. Studies have shown that 50% of all staff that leave your organization will take IP with them; 80% of those will knowingly use that IP at their new job.2 The primary reason behind these high numbers is a poor understanding by employees about IP and its importance. Since intellectual property is generated by your users, it follows that they should be tasked with identifying files that contain IP and the sensitivity. These actions will not only dramatically help your DLP systems protect IP from illicit access or sharing, but it will also remind users that this information has real value and belongs to the organization.

REASON 5 | ADDITIONAL BENEFITS OF CLASSIFICATION

Outside of enhancing DLP classification provides several other benefits that should not be overlooked.

Interoperability with the Entire Security Ecosystem

Persistent classification metadata offers the ability to trigger other protection systems based on classification, such as the automatic application of encryption like Ionic file protection, Microsoft AD Rights Management Services® (RMS) or S/MIME protection for email.

Data Retention Management

Classification simplifies data retention because it provides more information to a content archiving system and individual users to process when making decisions about the appropriate retention period. Classifications can include date or status fields that, when filled or edited, can instantly update the retention and disposition status.

Email Redactions

Email text can often contain sensitive information. By checking the email's classification level against the email content it is possible to alert users when they are about to send information that conflicts with policy. Users can be given the option to redact the sensitive data, replacing it with a black mark.

Flexible Email and Document Visual Markings

Classification can enable the application of customizable headers and footers, watermarks, email subject line marking, email message body labeling, dynamic disclaimers, and portion markings. These markings remind users of the information sensitivity which promotes responsible handling.

eDiscovery

Classification helps organizations avoid accidentally including too much of or even the wrong information in eDiscovery process. Classification labels can be used to help sort and qualify only the data required.

Insider Threat Detection

The effectiveness of insider threat detection improves significantly when it becomes possible to monitor how users interact with sensitive information. By providing identity to data there is no guesswork when analyzing exactly which files users are accessing, copying, and uploading. In addition, applying policy based on classification forces the malicious user to engage in activities that can quickly be flagged as suspicious, such as downgrading the classification of a file in order to bypass security protocols.

Start Protecting Your Data with Titus Classification

Protecting your data is a huge task that is made more difficult if data is not clearly identified. Although data loss prevention systems are extremely powerful and useful in the bid to keep private data private, the technology alone will not guarantee success. Finding the right balance between data protection policy and execution of that policy requires a thorough knowledge of your data. Lacking the ability to apply conclusive data identification, DLP systems can become a roadblock to business workflow.

For over 10 years, Titus has provided easy-to-use and highly scalable data classification solutions for enterprises in all industries. Titus Classification enables DLP and CASB systems to more effectively and efficiently protect data. Unlike purely automated solutions, our platform combines the benefits of automated classification with human insight. Our expertise and innovation in user-driven classification, combined with automated classification, enables our customers to engage users in data security and transform security culture. No other classification vendor provides the same level of flexibility and control in balancing different classification approaches. Many vendors offer basic classification capabilities. While Titus does basic classification very well, our extensive customer experience has shown even the most basic classification schemes come with sophisticated use cases.

Only Titus provides the policy granularity to support the current and future needs of large enterprises. One size does not fit all, which is why we give our customers the widest choice and flexibility in classification and labeling, policy enforcement, and DLP/CASB ecosystem integration.

Titus provides a complete data classification platform for data in use, in motion, and at rest. Our solutions identify and protect data wherever it resides, from desktop to mobile to cloud. Our platform was built to support both Cloud and on-premise environments so that customers have the flexibility to move to the Cloud when they are ready.

Titus customers include some of the largest and most successful organizations in the world. Leading banks, insurance companies, manufacturers, aerospace and defense contractors, energy companies, and government agencies choose Titus to classify their most sensitive data and augment the effectiveness of their DLP solutions. Our partnership with our customers provides us with unique insights into the challenges and rewards of data classification. We share these insights across our customer base through a highly effective deployment methodology and innovative products. No other classification vendor offers this level of experience and support to ensure customer success.

To find out how Titus can help your organization discover, classify, protect, and confidently share information, please visit www.titus.com.

Resources:

- *2013 Cost of Data Breach Study: Global Analysis.* Ponemon Institute, 2013. PDF.
- Brink, Derek. *DLP, The Ideal Referee*. Aberdeen Group, 2011. PDF.
- *Data Leakage Worldwide: Common Risks and Mistakes Employees Make.* Cisco, 2008. PDF.
- *Does Your Enterprise Classify Its Data?* Aberdeen Group, 2012. PDF.
- *Information Security Best Practices: Why Classification is Key*. Osterman Research, 2011. PDF.
- Kindervag, John, and Heidi Shey. *Strategy Deep Dive: Define Your Data*. Forrester Research, 2013. PDF.
- Kindervag, John. *Rethinking Data Loss Prevention with Forrester's DLP Maturity Grid*. Forrester Research, Inc., 2016. PDF.
- McMillan, Rob, and Eric Ouellet. *Best Practices for Data Loss Prevention: A Process, Not a Technology*. Gartner, 2012. PDF.
- Proctor, Paul. *Building an Effective Sensitive Data Classification and Handling Policy*. Gartner, 2010. PDF.
- Proctor, Paul. *Transform Your Security and Risk Program or Find Another Job.* Gartner, 2013. PDF.
- *The Role of Classification in Protecting Your Intellectual Property.* Aberdeen Group, 2012. PDF.
- *What's Yours is Mine: How Employees are Putting Your Intellectual Property at Risk.* Symantec, 2013. PDF.
- Shey, Heidi. Market Overview: *Data Loss Prevention*. Forrester Research, 2016. PDF

FORTRA

Fortra.com

About Fortra

Fortra is a cybersecurity company like no other. We're creating a simpler, stronger future for our customers. Our trusted experts and portfolio of integrated, scalable solutions bring balance and control to organizations around the world. We're the positive changemakers and your relentless ally to provide peace of mind through every step of your cybersecurity journey. Learn more at fortra.com.