FORTRA®

GUIDE (FORTRA)

Data Discovery And Classification: The Foundation Of Effective Information Governance

Introduction

Protecting data in accordance with its value or sensitivity is a critical part of information management and data governance. A data discovery and classification exercise, facilitated by the right technology tools, will enable an enterprise to find its sensitive data, label it with the appropriate sensitivity and ensure that you can demonstrate regulatory compliance by applying controls and policies consistently and accurately.

With a clear and accurate picture of the information that needs safeguarding you can design appropriate protection methods. The organisation must understand the data's value to the business – and also to hackers and competitors – as well as the impact if it were to be lost or leaked.

The volume of information held in, created by and flowing into organisations today makes data discovery a challenge. Every enterprise has millions of files, documents and messages to examine; often in many different data repositories. Alongside the relatively well-ordered structured data held in databases, companies have reams of unstructured data on user's desktops, which also store and share information via cloud applications or mobile devices. It is worth remembering that the value of a piece of data may also change over time for example, when the annual results are released they become public, which adds another layer of complexity.

In this paper you will learn the importance of why organisations should take a joint approach to tackling big data, coupling best of breed data discovery and data classification solutions as part of a data centric security approach.

Discover And Classify

Once it has been found the data can be classified according to its importance or sensitivity, using labels which determine how it should be handled and controlled. Data classification solutions attach the label as metadata as well as in a visual form. This allows it to be identified later, and makes security policies enforceable by ensuring that employees and security and information management solutions know exactly how it should be treated.

The process of building up a holistic view of your data involves three steps:

1. Determining how to identify the sensitive and business
2. critical data within the business.
3. Carrying out a discovery exercise to find out exactly
4. what legacy data exists and where it resides, as well as
5. what new data is being collected and created.
6. Determining who has access to all the data that has
7. been discovered, and how it is used.

Secure All Types Of Data

The process of finding and classifying data needs to encompass all enterprise data: the existing legacy information that is held within different repositories around the organisation, and new data that is produced or captured by the business on a daily basis. The unstructured data held by an organisation represents a particular challenge as it normally exists without the contextual wrapping found in a structured database.

Implementing a data discovery tool, in combination with a data classification solution, enables an organisation to cover all bases.

Locking Down Legacy Data

Data discovery tools will scan legacy data to locate sensitive information, so that the right classification can be applied. This is a major project that may precede the deployment of a classification solution, and might need the users to participate by manually validating some information. This activity provides a good opportunity to carry out other housekeeping tasks, for example, deduplication or deletion of redundant data.

Locating New Unclassified Material

Once a data classification solution has been implemented, discovery can become part of routine operational activity to locate new unclassified material, identify it and take action to classify it. A combination of user-driven and automated classification approaches is likely to be most effective. There will always be sources of unclassified information within the organisation, though over time the volume should get smaller and smaller. This may be data that is auto-generated internally, such as reports or exports of customer records, or which is coming into the business from external sources – such as emails, contracts or intellectual property (IP) from suppliers, or content that is downloaded by employees.

Supporting Search And Retrieval Tasks And Processes

The metadata 'tag' applied during classification increases the effectiveness of discovery tools in downstream security and governance activities. For example, in an eDiscovery scenario the metadata would enable tools to rapidly find all 'Confidential' data on public drives or quickly gather evidence for a legal case.

Discover And Monitor

Data is a living thing, and the ongoing security of a piece of information depends entirely on whether it is handled in accordance with its classification during its life. This means that data discovery needs to be a continuous activity, not a one-off project.

Once sensitive data has been identified and classified, it will have a metadata 'fingerprint' that enables the knowledge gained during the initial discover and classify phase to be used in different ways. The metadata allows each piece of sensitive data to be tracked through its journey, to see exactly what happens to it: who is accessing it and how often, what it is used for, and the systems, networks and repositories it flows through.

The visibility this delivers will enable potentially risky behaviour to be recognised and addressed, and will also support the monitoring and reporting requirements of compliance and data governance.

The Data Discovery Toolkit

Data discovery software and solutions make the process of identifying, classifying and monitoring sensitive and business critical data more efficient and effective. At a basic level, these examine file stores and databases, scanning for certain types of information, key words, criteria and classification metadata.

The most effective approach involves a scaled implementing of leading discovery tools as part of a security "ecosystem" that matches best-of-breed solutions to your requirements.

The ecosystem should be centred on a flexible and powerful data classification platform that allows the organisation to select tools on a cost and functionality basis, and remove and upgrade tools in the ecosystem as their needs evolve.

As a standalone solution, Fortra's knowledge-based classification tools ensure that data at rest is correctly classified.

Fortra uses AI techniques to take information from different sources in order to identify and classify data files at rest. Input data can either be from Fortra, content and concept, file scanning or third-party discovery and search tools.

- Enhanced classification of data at rest based on content and context
- Support for all Windows mountable drives
- SharePoint (local and online)
- OneDrive for Business
- Files classified according to organisation's group policy using AI
- Complete audit of every files content and metadata for data inventory
- Helps identify sensitive content (PII, PCI DSS etc.) as well as intellectual property, financial and regulated data

Fortra's Data Classification Suite also works alongside best-of-breed discovery tools from third-parties to provide a complete data management and protection solution.

There are a number of discovery tools on the market, the majority of which have the capability to:

- Identify the key categories of information. This could be as simple as pinpointing keywords or expressions that characterise a type of data – for instance Personal Healthcare Information (PHI).
- Find all the data matching those categories.
- Apply 'remediation' actions – for example, relocate it, classify it or protect it.

Some offer extended features that support the protection and governance of data that has been discovered, some of which use artificial intelligence (AI) techniques to make the process more efficient. These could include:

- Monitoring the disposition and usage of classified information to enable any activity that violates security policy to be detected – for example, where confidential board-level documents are being accessed by people who are not board members.
- Analysing user behaviour around classified data to enhance decision making.
- Automation of data retention policies – using rules to find and migrate, archive, or delete files.
- Modification of access permissions.

Some vendors have developed innovative solutions with enhanced functionality that will strengthen the discovery and classification process if they are integrated into the security ecosystem. For instance, software that further leverages classification metadata to manage, analyse, and secure enterprise data will provide the business with an accurate 'map' of its information landscape, identify when sensitive data has been exposed and automatically safeguard it.

The data loss prevention (DLP) solutions that many organisations are already using may incorporate specialist discovery tools or file analysis products, so it is worth investigating whether you already have these at your disposal before researching the market and approaching vendors.

Once you have found your most sensitive and business critical data, determined the potential risks to its security, prioritised it and classified it, you can make informed decisions about how to protect it using controls such as DLP solutions, encryption, access control, policy and procedures.

The Human Element In Classification

Automated technology solutions can provide support with the task of identifying valuable or sensitive information, and also enforcing controls downstream. However, the insight of data owners and users is vital to ensure that information is categorised correctly and so the discovery and classification exercise should always be a balance between automated and manual processes.

Humans are best-placed to put a value on the data they store, create or handle. They can apply their knowledge of the context around it to determine what is most important, what is redundant and can be deleted or archived, the potential impact if the data is leaked or lost and, ultimately, the most appropriate classification to apply.

User-driven classification solutions empower users to attach a visual and metadata label to data at the point of creation or manipulation. The metadata tag directs the actions of downstream security and information management toolsets, safely locking down the data through its journey and also closing the discovery loop – making it easier to locate, retrieve and monitor later as part of data governance.

Evolving Classification

Classification schemes may have a simple basis – beginning with a few levels, such as ‘internal only’, ‘commercial in confidence’ or ‘public’. This root will need to evolve over time however as the organisation matures, to include more granular business-specific and compliance-specific sub levels. Labels should reflect all the different ways the business needs to ‘slice and dice’ its data, such as those relating to the activity of specific business units, or compliance with US export control and ITAR for example.

The selected classification solution should therefore be one that can scale up to meet evolving needs. Organisations implementing a classification solution 5 years ago weren’t aware of the requirements GDPR compliance would bring to their organisation - something which is a forefront requirement today. Likewise the companion discovery tools need the flexibility to track that evolution.

Common Pitfalls

Combining data discovery with data classification processes and solutions can have significant benefits as previously outlined, however, there are common challenges that organisations face when tackling a data at rest project. Some of these can be summarised below:

- **Lack of data owner engagement** - as part of the discovery and classification process it is necessary to involve data owners to evaluate and determine whether specific data sets should be classified as sensitive. Given the scale of the task for some organisations this may be perceived as onerous, however, the benefits of going through this process far outweigh the negatives. Awareness and communication around the project goals helps to explain the importance of the project to the owner and the organisation.
- **Scale of data discovery task** - given the nature of big data, many organisations become overwhelmed with the task from the outset. The key here is to break the project down into stages, and pace the project according to available resources.
- **Matching the classification approach with the business needs and resources** - for classifying data at rest the level of discrimination needed depends on balancing the resources available with the sensitivity of the data you are dealing with. For some organisations it might be more appropriate to perform a relatively quick discovery scan for specific file types or keywords, marking those as appropriate, and then blanket classify the remaining data with a common label to ensure the most sensitive data is tagged and does not get inadvertently leaked.

Conclusion

With the emergence of new data protection regulations, organisations must take a combined data discovery and data classification approach to ensure sensitive data is adequately protected. Combining best-of-breed data discovery and data classification technologies provides organisations with the right tools for the job, and importantly enhances functionality that strengthens the discovery and classification process, particularly if they are integrated into a 'metadata' security ecosystem. Once you have found your most sensitive and business critical data, determined the potential risks to its security, prioritised it and classified it, you can make informed decisions about how to protect it using controls such as DLP solutions, utilising encryption or other post-delivery controls.

In today's digital economy, providing the right tools to tackle big data protection issues is paramount to ensuring success and safeguards against financial and business risk associated with data loss incidents. In short, combining best-of-breed solutions for discovery and classification enables organisations to cover all bases – from classifying data being created and in transit, to critically classifying sensitive data at rest.

Global brands trust us to protect their sensitive data:

pwc

Ministry of Defence

Honeywell

PRUDENTIAL

Department of Defense United States of America

AMGEN

SOCIETE GENERALE

Allianz

Raytheon

More Information

For more information about how you can implement a data classification solution as part of your data protection strategy, please visit fortra.com.

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.