

# THE DEFINITIVE GUIDE TO DATA CLASSIFICATION

DATA CLASSIFICATION FOR DATA  
PROTECTION SUCCESS



# TABLE OF CONTENTS

<b>03</b>	Introduction
<b>04</b>	Part One: What is Data Classification?
<b>06</b>	Part Two: Data Classification Myths
<b>08</b>	Part Three: Why Data Classification is Foundational
<b>12</b>	Part Four: The Resurgence of Data Classification
<b>16</b>	Part Five: How Do You Want to Classify Your Data?
<b>21</b>	Part Six: Selling Data Classification to the Business
<b>27</b>	Part Seven: Getting Successful with Data Classification
<b>33</b>	Part Eight: HelpSystems Data Classification & Protection

# WHY READ THIS GUIDE?

## THERE ARE TWO TYPES OF COMPANIES: THOSE THAT RUN ON DATA AND THOSE THAT WILL RUN ON DATA

InfoSec professionals will perennially be challenged with more to do than time, budget, and staffing will allow. The most effective method to address this is through prioritization, and in the case of your growing data, prioritization comes from data classification. In this guide you will learn what classification is, why it is important, even foundational to data security, and much more.

## HOW TO USE THIS GUIDE

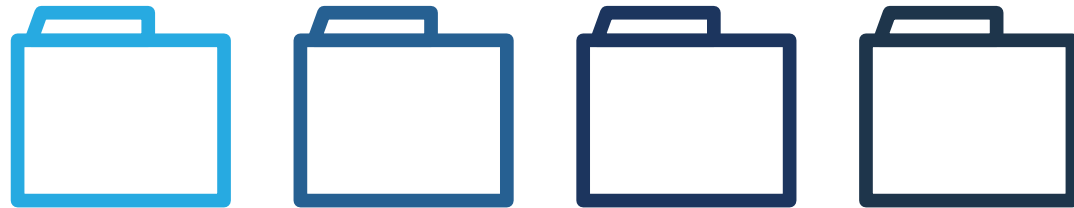
IF YOU ARE...	GO TO...
<b>New to data classification</b>	Part One: What is Data Classification
<b>Learning how data classification drives your data security strategy</b>	Part Three: Why Data Classification is Foundational
<b>Trying to understand the different classification methods</b>	Part Five: How Do You Want to Classify Your Data
<b>In need of speaking points for building internal support</b>	Part Six: Selling Data Classification to the Business

# PART ONE

# WHAT IS DATA CLASSIFICATION?

# DATA CLASSIFICATION

**WHAT:** Data classification is the process of consistently categorizing data, using visual and metadata labels, based on specific and pre-defined criteria so that data can be efficiently and appropriately protected.



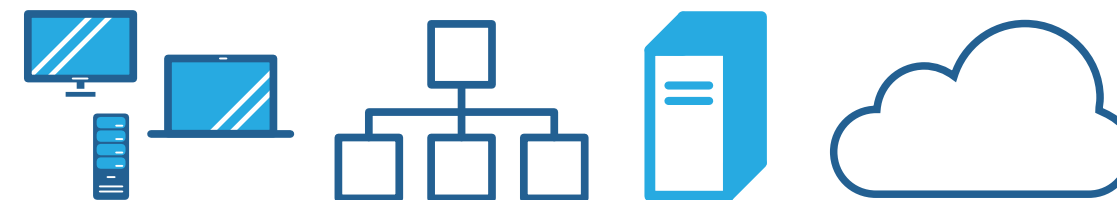
**WHY:** The need for classification can be driven by governance, company compliance, regulatory requirements (GDPR, HIPAA, PCI, CCPA and more), protection of intellectual property (IP), or perhaps most importantly, by the need to simplify your security strategy (more about that later).

**HOW:** There are a few key questions organizations need to ask to help define classification categories. Answering these will guide your data classification efforts and get the program started.

- What are the data types? (structured vs unstructured)
- What data needs to be classified?
- Where is the sensitive data located?
- What are some examples of classification levels?
- How can data be protected and which controls should be used?
- Who has access to what data?

## BEFORE YOU CAN CLASSIFY

Data discovery is closely aligned with classification; before you can classify data you need to know what you have. Data discovery needs to look at the endpoint, on network shares, in databases, and in the cloud.



PART TWO  
**DATA  
CLASSIFICATION  
MYTHS**

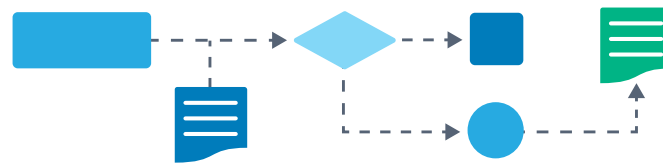
# 3 MYTHS OF DATA CLASSIFICATION



## MYTH 1: LONG TIME TO VALUE.

Automated classification drives insights from day one. Automation for both context and content brings order to all your sensitive data; quickly and easily.

Data collection and visibility can continue until the organization is prepared to deploy and operationalize a policy. Even without a policy, insights from automated data classification can drive security improvements.



## MYTH 2: IT'S TOO COMPLICATED.

Many data classification projects get bogged down by starting with overly complex classification schemes. When it comes to classification more is not necessarily better; more may just be more complex.

PricewaterhouseCoopers, Forrester, and AWS all recommend starting with just three categories. Starting with a simplified classification policy helps to get your program off the ground. If after deployment more granular levels are needed, your decision will be driven by data, or regulatory requirements, not simply speculation.



## MYTH 3: IT'S ANOTHER LEVEL OF BUREAUCRACY.

Data classification can be an enabler and a way to simplify data protection. By understanding what portion of your data is sensitive, resources are allocated appropriately.

Users understand what needs to be protected. Sensitive and regulated data is prioritized; public data is given lower priority, or destroyed, to eliminate future risk to its theft.

PART THREE

# WHY DATA CLASSIFICATION IS FOUNDATIONAL

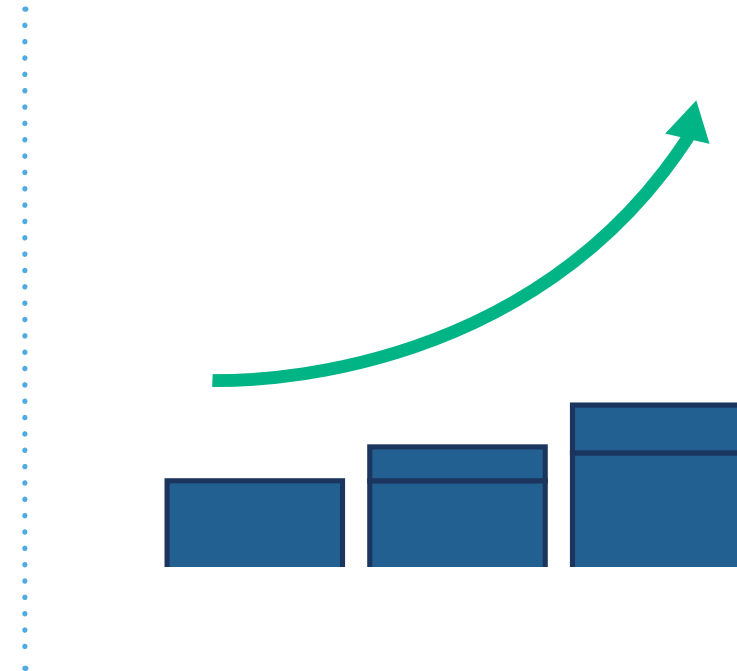


# IT'S EASIER TO MANAGE THE DATA DELUGE WITH CLASSIFICATION

Organizations generate huge volumes of data. This comes as no surprise, but what might be surprising is the accelerating volume at which the data is being created. As an InfoSec professional responsible for protecting digital data, you're going to need a new approach to stay ahead of the data deluge.

Classification enables you to:

- Avoid taking a "one size fits all" approach (inefficient!)
- Avoid arbitrarily choosing what data to expend resources protecting (risky!)



IDC estimates that the digital universe is growing at ~40% year over year.

(source: A Day in Data. IDC/Raconteur)

# WHY GARTNER THINKS DATA CLASSIFICATION IS FOUNDATIONAL

"To implement an effective data classification program, security and risk management leaders tasked with data security must establish a data classification program by shifting focus from user awareness and training toward automation and the enrichment tools that generate metadata."

"Data classification is vital as it is useful in supporting controls for data security and governance such as data loss prevention (DLP), data access governance and enterprise digital rights management (EDRM)."

The Gartner logo is displayed in a large, bold, dark blue font. A vertical dotted line is positioned to the left of the logo, extending from the top of the text area down to the bottom of the slide.

# WHY FORRESTER THINKS DATA CLASSIFICATION IS FOUNDATIONAL

## START FROM DATA CLASSIFICATION

“Security & Risk (S&R) professionals must start from data classification to build their data protection strategy.”

## UNDERSTANDING AND KNOWING YOUR DATA IS THE FOUNDATION

“For many S&R pros, data security initiatives quickly zoom in on controlling access to data or encrypting data. But many overlook that understanding and knowing your data is the foundation for both data security and privacy...”

## IF YOU DON'T KNOW WHAT YOU HAVE, YOU CAN'T PROTECT IT

“If you don't know what you have [data], where it is, and why you have it, you can't expect to apply the appropriate policies and controls to protect it.”



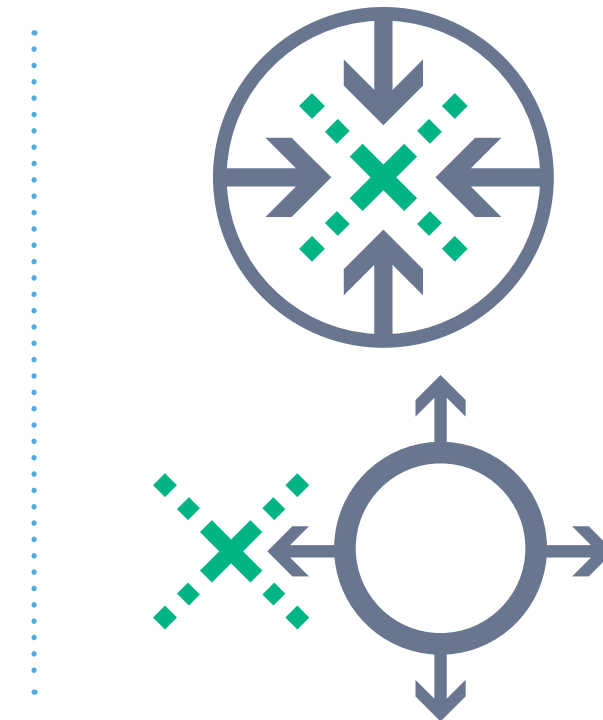
PART FOUR  
**THE RESURGENCE  
OF DATA  
CLASSIFICATION**

# CLASSIFICATION HELPS PROTECT AGAINST ALL THREATS

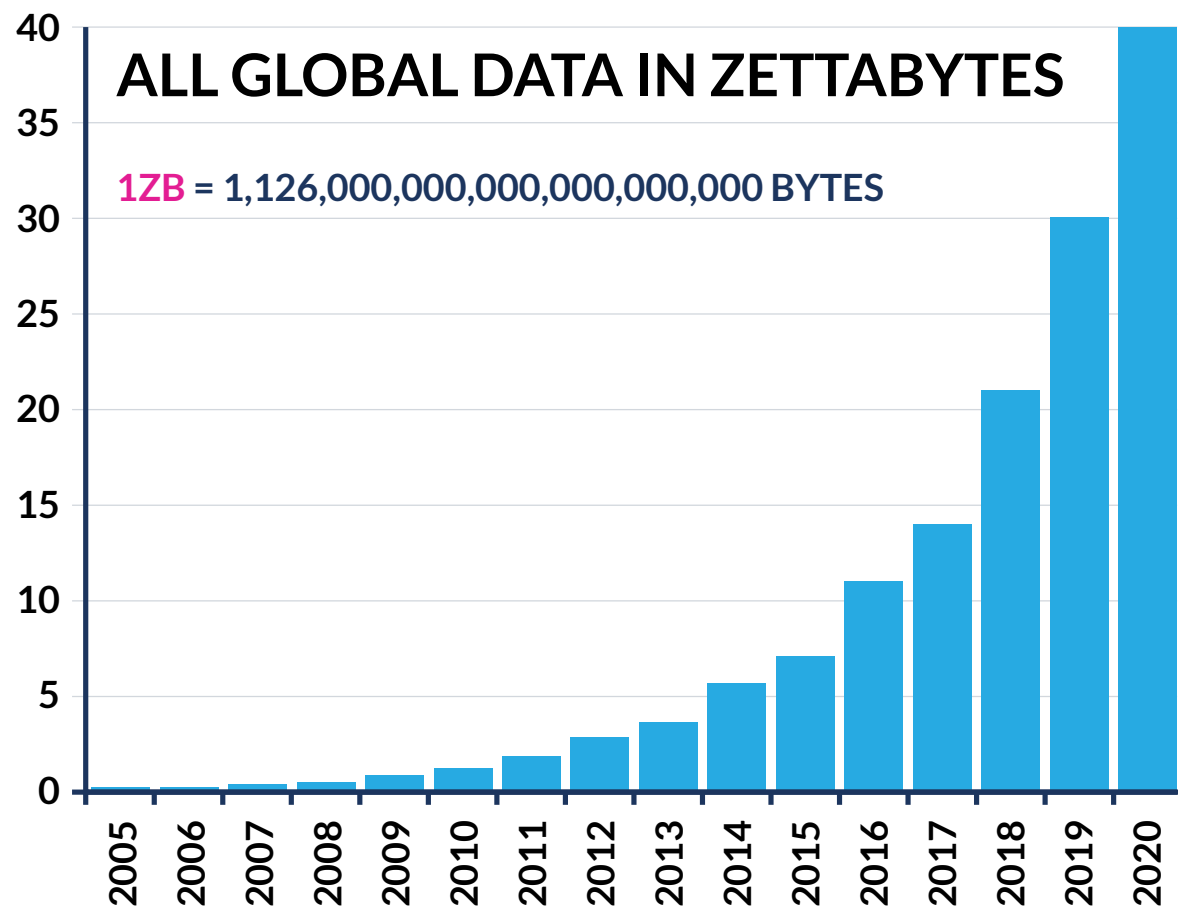
The value to classification was once limited to protection from insider threats. With the growth in outsider threats, classification takes on a new importance. It provides the guidance for information security professionals to allocate resources towards defending the crown jewels against all threats.

Internal actors cause both malicious and unintentional data loss. With a classification program in place, the mistyped email address in a message with sensitive data is flagged. Files that are intentionally being leaked are classified as sensitive and get the attention of security solutions, such as Data Loss Prevention (DLP).

External actors often seek data that can be monetized. Understanding which data within your organization has the greatest value, and the greatest risk for theft, is where classification delivers value. By understanding the greater potential impact of an attack on sensitive data, advanced threat detection tools escalate alarms accordingly to allow more immediate response.

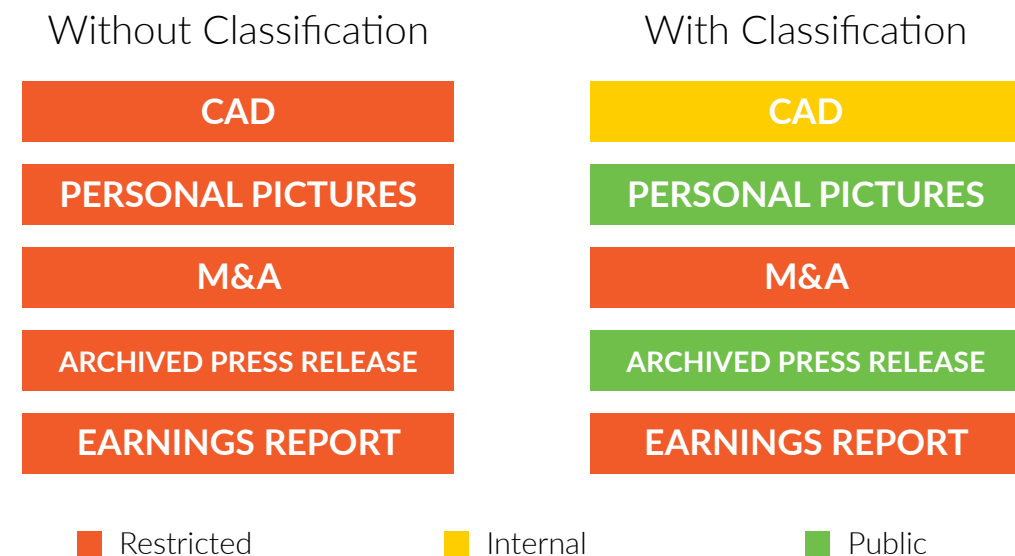


# BIG DATA IS DRIVING BIG CLASSIFICATION NEEDS



## SOMEWHERE IN YOUR DATA DELUGE IS:

- A **CAD** drawing of the next generation iPhone
- **Personal pictures**
- **M&A** plans
- **An archived press release** announcing your previous acquisition
- A quarterly **earnings report** in advance of reporting date



# ADOPTION MOMENTUM

"72%

of security decision makers surveyed said that they are implementing, have implemented, or are expanding/upgrading implementation of data classification."

Just having a classification solution isn't always enough, read on to learn how to align classification to your business needs.

## PART FIVE

# HOW DO YOU WANT TO CLASSIFY YOUR DATA?



# ONE SIZE DOES NOT FIT ALL

## CHOOSE CLASSIFICATION METHODS BASED ON THE DATA TYPES MOST IMPORTANT TO YOUR BUSINESS

Combine privacy regulation adherence efforts with the security classification initiatives. As information can be categorized by nature or by type. Regardless, records should also be classified by risk categories as to indicate the need for confidentiality, integrity and availability.

The Gartner logo, consisting of the word "Gartner" in a bold, dark blue sans-serif font, followed by a registered trademark symbol (®). The logo is positioned in the bottom right corner of the slide, partially overlapping a light green abstract graphic element.

# DATA CLASSIFICATION METHODS

**Content-based** classification inspects and interprets files looking for sensitive information. Methods include fingerprinting and regular expression.

**Content-based** answers “What is in the document?”



**Context-based** classification looks at application, location, or creator among other variables as indicators of sensitive information.

**Context-based** answers “How is the data being used?” “Who is accessing it?” “Where are they moving it?” “When are they accessing it?”

**User-driven** classification relies on manual, end-user selection.

**User-driven** relies on user knowledge and discretion at creation, edit, or review to identify sensitive documents.

# WHICH CLASSIFICATION METHOD?

The decision around which classification method to use is usually a question of which to start with as opposed to picking just one. Each provides insight; combining them provides greater security. Including context and content with a user-driven approach delivers the backstop needed to mitigate the impact of misclassified data (either unintentionally or maliciously).



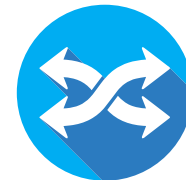
## COMPLIANCE

Compliance data is often structured and/or residing in predictable locations. Leading with a content-based classification will provide the greatest ability to accurately classify PII, PHI, PCI, and GDPR data.



## IP PROTECTION

Intellectual property seldom follows a pattern like a credit card number. To address, this context classification looks to other attributes to assign classification. The application used or the storage location are two ways IP can be classified to support data protection.



## MIXED ENVIRONMENT

Where a mix of regulated data and intellectual property drive enterprise growth, organizations looking to better understand and protect their data look to a blended approach.



## USERS

Data owners should know their data best. A user-based classification approach allows them to apply this knowledge to improve classification accuracy.

# COLLABORATE AND COMBINE FOR SUCCESS

## GARTNER RECOMMENDATIONS

- To identify, tag and store all of an organization's data, SRM leaders and chief data officers (CDOs) should collaboratively architect and use classification capabilities.
- Implement data classification as part of a data governance program.
- Use a combination of user-driven and automated data classification.

The Gartner logo is displayed in a bold, dark blue, sans-serif font. The word "Gartner" is followed by a registered trademark symbol (®).

*(source: Gartner Hype Cycle for Cyber and IT Risk Management, 19 July 2021)*

PART SIX

**SELLING DATA  
CLASSIFICATION TO  
THE BUSINESS**

# DATA CLASSIFICATION TEAM

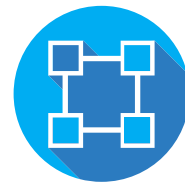
Data classification decisions can impact all employees. Who are the players within your organization you need to talk to and what do you communicate to them?



## CIO & CISO

The ultimate technical responsibility for data protection falls upon one, or both, of these roles. Where the CIO is running the IT operations, the CISO is securing the IT operations. For them to be effective they both need to understand the sensitive data landscape.

- **CIO:** Classification guides and simplifies IT infrastructure investment decisions by cataloging volume, location, and type of sensitive data.
- **CISO:** Classification highlights where to allocate the security resources and can spot security gaps before they become breaches.



## BUSINESS UNIT LEADERS

The P&L leaders who watch the top (and bottom) line numbers of the business units. This role has a more immediate reason to support data classification – loss of data in their business unit could result in revenue impact, fines, or both.

Classification drives visibility and protection of both customer data (PII) and the product development data (IP) that fuels growth.



## DATA CREATORS

The feet on the street; the knowledge workers that are often writing the code, creating the CAD documents, or drafting the M&A proposals. They are closest to the data and are instrumental to any protection program, it must serve its protective purpose without impeding business.

Including the users in a classification program heightens awareness to the need to protect data and the negative repercussions if that data leaks.



## LEGAL/COMPLIANCE

Legal is there when things go wrong and data leaks. Often the backstop in a data protection program, legal needs to understand the scope of the sensitive data (exposure) and the protection in place (mitigating factors) to ensure the organization is properly managing the risk. Risk is unavoidable in business, but which risks to accept needs to be a calculated and conscious decision.

# CLASSIFICATION "QUICK WINS"

## TIP

Get users involved early. Any change that requires workflow modifications can be a source of friction. If your data classification project involves user-driven classification (and not all do, some rely wholly on automated data classification techniques), getting the users on board ahead of the project means that when roll-out happens they are educated, enabled, and understand the needs, along with the benefits, in **\*their\*** terms.



# POSITIONING DATA CLASSIFICATION

## DATA CHAMPIONS

The data champions are those who have the most invested in the data. The goal here is to ensure they understand:

- What they are creating has value
- The value is worth protecting from both internal and external threats
- They are an important piece of the protection

## EXECUTIVES

To a data intensive organization (something that most are becoming whether they realize it or not) protecting their data is paramount to sustainable competitive advantage. They need to understand:

- Classification can drive revenue growth by enabling secure partnerships and growth initiatives
- Classification can reduce spend by limiting the scope of data needing protection and increasing the efficiency of existing investments
- Classification can reduce risk by highlighting where sensitive data is and where it is going






# OVERCOMING OBJECTIONS

“We’ve gotten along just fine without it.” This passive message is akin to saying “I’ve never needed insurance in the past,” and reflects a misunderstanding of the importance of classification or a misperception that it is only for more mature organizations. While organizations can protect their data without classification, it comes at the expense of efficiency.

- **With** classification, data loss prevention and advanced threat protection have the insight to understand the difference between regulated, internal only, and public data. This insight intelligently elevates data risks based on the impact of a breach.
- **Without** classification, data protection solutions, including data loss prevention and advanced threat protection, will be prone to higher false positives and false negatives, and alerts will be of lower fidelity.


Building your data protection strategy on classification is the foundation needed for success.

# MORE JUSTIFICATION FOR CLASSIFICATION



“Data classification enables an effective and efficient prioritization for data governance programs that span value, security, access, usage, privacy, storage, ethics, quality and retention.”

“It is vital to security, privacy and data governance programs. It also allows organizations to have the required knowledge about the sensitivity of the data they process.”



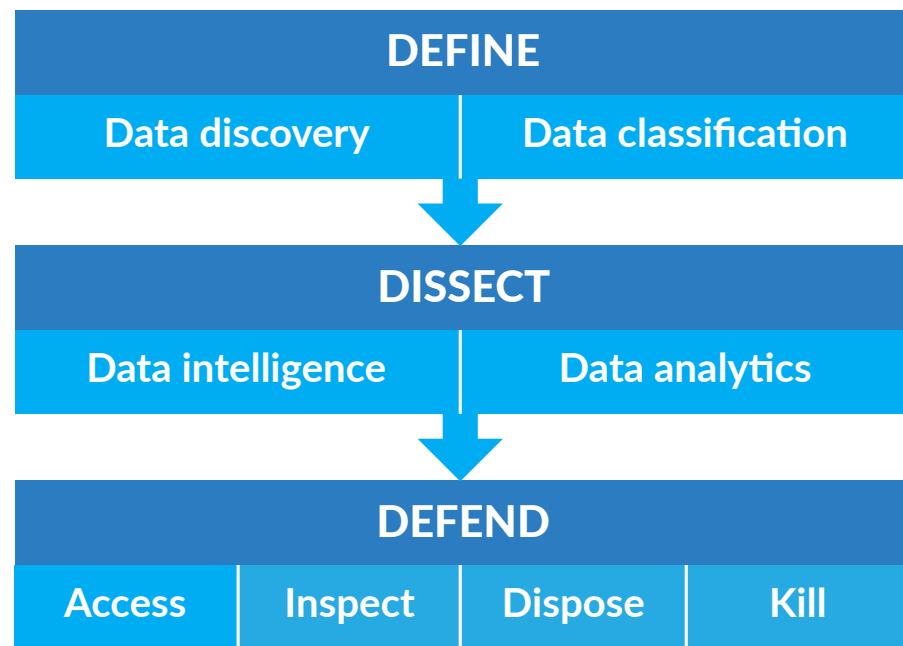
**Gartner**<sup>®</sup>

*(source: Gartner Hype Cycle for Cyber and IT Risk Management, 19 July 2021)*

PART SEVEN  
**GETTING SUCCESSFUL  
WITH DATA  
CLASSIFICATION**

# DATA PROTECTION FRAMEWORK

Many organizations need help getting started. Forrester created a framework to guide you on this journey. Their “Data Security & Control Framework” (figure below) breaks the problem of controlling and securing data into three steps: Define, Dissect, Defend. With these steps completed your organizations better understands your data and can then allocate resources to more efficiently protect critical assets. At the top of their framework: Discovery and Classification.



**DEFINE:** This involves data discovery and data classification.

**DISSECT:** This involves data intelligence (extracting information about the data from the data, and using that information to protect the data) and data analytics (analyzing data in near real-time to protect proactively toxic data).

**DEFEND:** To defend your data, there are only four levers you can pull – controlling access, inspecting data usage patterns for abuse, disposing of data when the organization no longer needs it or “killing” data via encryption to devalue it in the event that it is stolen.



(source: A Strategic Guide For Controlling And Securing Your Data, Forrester’s Data Security Control Framework by Heidi Shey January 19, 2021)

# DATA CLASSIFICATION PROCESS

You've bought in on a data classification program, what are the key elements to drive success?



# YOUR CLASSIFICATION GUIDELINE

To be effective, your classification program needs a well defined policy. This includes the right number of categories and clear mapping of your data to those categories. PricewaterhouseCoopers, Forrester, and AWS, among many security analysts and consultants, recommends you start with just three categories: Public, Private, and Restricted. While simple classification might be where you start, it's highly likely it won't be where you end up. From ever-increasing global data protection regulations, to adaptations for different end-user communities, it is important to implement a solution that will allow for this future flexibility.

Below is an example policy matrix illustrating the document types, risks, and protective controls. ([Click here for a blank template](#))

	PUBLIC	PRIVATE	RESTRICTED
DEFINITION	Documents are acceptable for public use without restrictions.	Documents are not to be distributed externally unless under specific conditions.	Documents are subject to compliance restrictions (PCI, HIPAA) and are not to be distributed externally unless under specific conditions.
EXAMPLE DOCUMENT	Product datasheet, job postings.	Strategic planning document, product roadmaps, CAD drawings.	Customer database, payment card information, health record information.
REPERCUSSIONS IF LEAKED	None	Loss of competitive advantage, loss of brand equity, reputational damage.	Fines, customer churn, reputational damage.
CONTROLS IN PLACE	N/A	Education and awareness training, file encryption, data loss prevention, advanced threat protection, reporting and auditing.	Education and awareness training, automated encryption, data loss prevention, advanced threat protection, reporting and auditing.

# YOUR CLASSIFICATION TEMPLATE

	PUBLIC	PRIVATE	RESTRICTED
DEFINITION			
EXAMPLE DOCUMENT			
REPERCUSSIONS IF LEAKED			
CONTROLS IN PLACE			

# DATA CLASSIFICATION GUIDANCE - START OFF SIMPLE!

Table 1: Three Categories of Data, Classified by Risk

Data Type	Description
Public	This is data published on a publicly facing website or in other official external communications, such as social media feeds and marketing collateral.
Internal	This data, for internal use only, appears in routine business communications and documents created as part of normal, day-to-day activities. It includes data in the majority of internal emails.
Confidential	This is sensitive data that typically requires special handling procedures. It can be data subject to regulations, intellectual property, or information that is not publicly known or available internally, such as merger and acquisition documents, corporate financial reports and HR data.

## Resist the Urge to Expand the Classification Schema Without Good Reasons

*“There is no standard classification schema as datasets and appetites for risk vary greatly across organizations. Many successful deployments of data classification programs by organizations focused on regulatory compliance or intellectual property use a variation of the simple three-classification approach to grouping data according to risk.”*



PART EIGHT

**HELPSYSTEMS**

**DATA CLASSIFICATION  
& PROTECTION**

# HELPSYSTEMS DATA PROTECTION

To protect your expanding and valuable pool of data from insider and outsider threats organizations need a data-centric plan. Below is a 4 step framework to take control of, and protect, your knowledge assets and keep them protected without impacting the speed of business.

**Discovery** - You need to know exactly where your sensitive data is to protect it. This includes on laptops, desktops, and servers, but also in the cloud.

**Education & Enforcement** - Provide real-time alerts for potentially risky behavior allowing users to self correct. If needed, implement data protection policies and ensure they are followed.



**Classification** - Structure and organization for your data enables your data security program and delivers more accurate protection.

**Policies** - Now that you know the Where and the What, it is time to define How you are going to protect it.

# HELPSYSTEMS DATA CLASSIFICATION

HelpSystems data classification solutions enable classification via context, content, and user-based methods to cover the spectrum from fully automated to fully manual classification.

Our data classification solutions, Titus and Boldon James, integrate into our full data protection suite offering, including DLP from Digital Guardian and DRM from Vera. This integration, and the built-in automation, delivers a more accurate data protection program to limit false positives and false negatives.

By combining data discovery, data classification, policies, and enforcement, HelpSystems data classification solutions provide the comprehensive data protection needed to stop data theft.

## EXAMPLE METHODS



### CONTENT

A database is scanned, and PCI regulated data is discovered and analyzed. Any outbound message is compared with this database fingerprint for a match. If found, the message can be encrypted, quarantined, blocked, or logged.



### USER

A new project requires creation of multiple CAD files. At “save,” the data owner self-selects these to be classified as “sensitive” intellectual property. When an outbound communication contains these documents the message can be encrypted, quarantined, blocked, or logged.



### CONTEXT

Detailed seismic studies of a newly-relevant region for petroleum exploration are stored on a designated server and created using a specialized application. Context based classification sees files location and application used; any message meeting specific file and application criteria can be automatically encrypted, quarantined, blocked, or logged.

# AUTOMATION CONTINUUM

Automation drives repeatability and predictability, it also speeds implementation time. But it needs to be augmented with the knowledge of the data owners. HelpSystems delivers classification options that cover the spectrum from fully automated to fully user-driven to match your organizations' needs.

- Automated context and content classification gets your program operational quickly and provides consistent results for more accurate data security and to demonstrate compliance.
- User-driven classification incorporates the intimate knowledge and bigger-picture view data owners possess, delivering the accuracy and compliance automation and AI cannot (yet).
- A blend of user-driven and automated provides the insights needed to scale securely and protect all your sensitive data.

## Fully automated



Most DLP solutions require you to spend time identifying and classifying your sensitive data before protection starts. Upon installation, HelpSystems data classification proactively finds, classifies, and tags files.

## Partially automated



Classify and tag based on predefined **context**, such as file properties, file location, or application used.

Classify and tag based on predefined **content**. Content inspection engine identifies patterns in files or databases then applies classification tags to them.

## Fully user-driven



**User-driven** classification relies on the data owner to apply the tag to the document at creation, or after modification.

# LEADING DATA PROTECTION FROM HELPSYSTEMS

Organizations rely on sensitive data to serve their customers or patients, fuel innovation, and grow. Security leaders need a way to find and understand that data, then protect it from loss or theft while within their extended enterprise, and securely share it outside of their extended enterprise. HelpSystems leading data protection offering combines data classification from Boldon James and Titus, with data loss prevention from Digital Guardian, and digital rights management from Vera to deliver data protection throughout the entire data lifecycle.

- Data Discovery
- Data Classification
- Data Loss Prevention
- Managed Detection & Response
- Digital Rights Management
- Analytics
- Reporting
- System Management

**boldonjames**  
by HelpSystems

**titus**  
by HelpSystems

The logo for Digital Guardian, featuring a stylized graphic of two overlapping shapes, one grey and one pink, above the text "DIGITAL GUARDIAN" in a bold, sans-serif font.  
**DIGITAL GUARDIAN**  
by HelpSystems

**vera**  
by HelpSystems

# THE DEFINITIVE GUIDE TO DATA CLASSIFICATION

2022 EDITION

QUESTIONS?

U.S. 800.328.1000

Outside U.S. +44(0)870 120 3148

info@helpsystems.com

www.helpsystems.com



©2022 HelpSystems. All rights reserved.

