FORTRA

The Definitive Guide To Data Classification

Data Classification For Data Protection Success

2022 Edition

Table Of Contents

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

Part One

What Is Data Classification?

Data Classification

What: Data classification is a process of consistently categorizing data based on specific and pre-defined criteria so that this data can be efficiently and effectively protected.

Why: Classification can be driven by governance, company compliance, regulation (HIPAA, PCI, or CCPA), protection of intellectual property (IP), or perhaps most importantly, by the need to simplify your security strategy (more about that later).

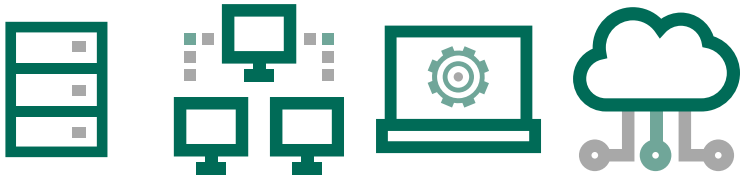
HIPAA PCI CALIFORNIA CONSUMER PRIVACY ACT

How: There are a few key questions organizations need to ask to help define classification buckets. 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 have to find it though. 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 because of overly complex classification schemes. When it comes to classification more is not better; more is just more complex.

PricewatershouseCoopers, Forrester, and AWS all recommend startting with just three categories. Starting with three can dramatically simplify getting your program off the ground. If after deployment more are needed your decision will be driven by data, not 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.

Everyone understands 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 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!).

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

IDC
Analyze the Future

"By 2025 463 exabytes ofdata will be created every single day. That's the equivalent of 650 years of 8K quality video."

Why Gartner Thinks Data Classification Is Foundational

Document The Crown Jewels Of The Organization
"Identify your organization's 'crown jewels' — information and services that are critical to meeting strategic business objectives — and tailor technical and procedural controls to balance protection and business operating realities."

Focus On Foundational Controls
"Focus on controls that broadly address the problem, such as implementing people-centric security and data classification. These controls are the foundation upon which additional controls can be built."

Use Data Classification As An Enabler
"In effect, data classification enables a less restricted handling of most data by bringing clarity to the items requiring the elevated control."

Gartner®

(source: How to Succeed With Data Classification Using Modern Approaches, Published 25 March 2022, Ravisha Chugh, Bart Willemsen, Nader Henein)

Why Forrester Thinks It's Foundational

Start From Data Classification

"Security & Risk 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."

FORRESTER®

(source: Rethinking Data Discovery and Data Classification Strategies, Forrester Research Inc., July 10, 2018, Heidi Shey)

Part Four

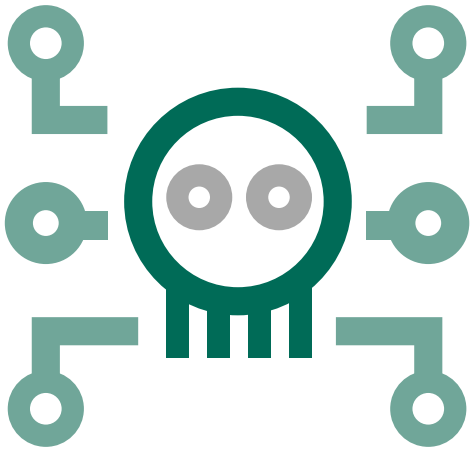
The Resurgence Of Data Classification

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 pros 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).

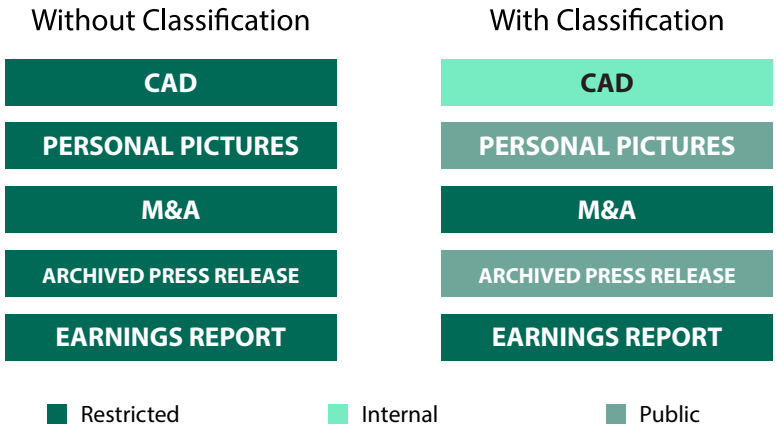
External actors seek data that can be monetized. Understanding which data within your organization has the greatest value, and the greatest risk for theft, is where classification delivers value. By understanding the greater potential impact of an attack on sensitive data, advanced threat detection tools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

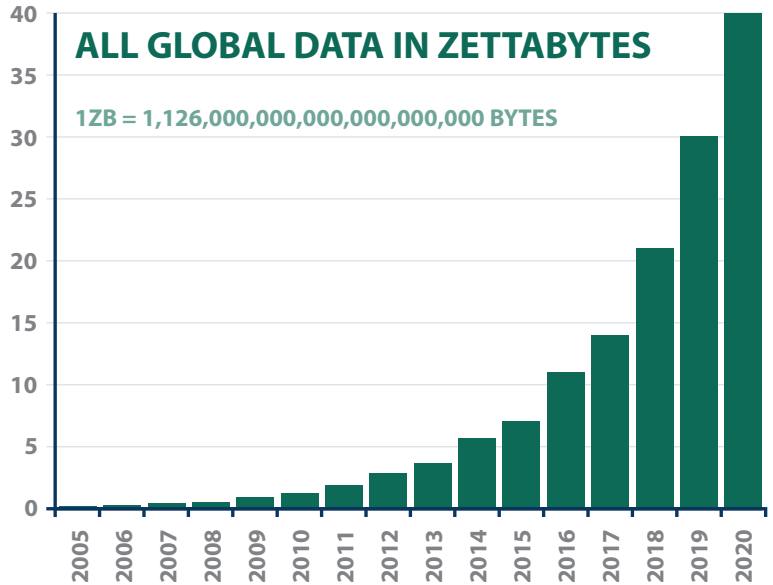


Big Data Is Driving Big Classification Needs

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

44% of enterprises currently use data classification as part of their overall information risk program.

33% more are evaluating data classification with over half that number planning to implement data classification in the next 18 months

(source: Forrester's Global Business Technographics® Security Survey, Forrester Research Inc., 2015)

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

"Use the combination of data classification approaches and techniques most appropriate for the datasets they're trying to secure."

Gartner®

(source: Innovation Insight for Unstructured Data Classification, Gartner, May 2017, Marc-Antoine Meunier, Brian Reed)

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-based classification relies on manual, end-user selection.

User-based relies on user knowledge and discretion at creation, edit, or review to flag 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-based 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.

Automated And Manual — Better Together

Dynamic Data Classification Requires Both Tools And Human Intervention

"Recognize that data is a living thing. Dynamic data classification requires the integration of both manual processes involving employees as well as tools for automation and enforcement. Human intervention provides much-needed context for data classification, while tools enable efficiency and policy enforcement."

(source: Rethinking Data Discovery and Data Classification Strategies, Forrester Research Inc., July 10, 2018, Heidi Shey)"

Part Six

Selling Data Classification To The Business

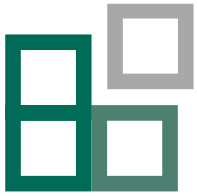
Data Classification Team

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

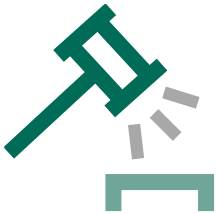
The P&L leaders who watch the top (and bottom) line numbers of the business units. This role has a more immediate reason to support data classification – loss of data in their business unit could result in revenue impact, fines, or both.

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

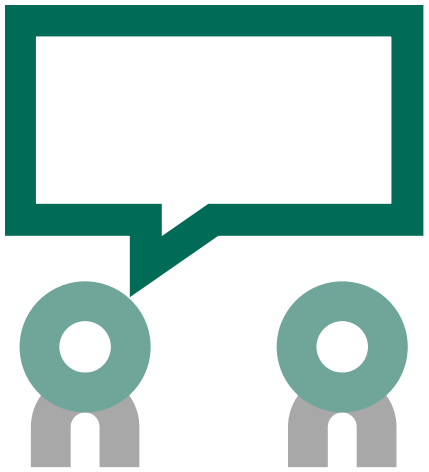
Get users involved early. Any change that requires workflow modifications can be a source of friction. If your data classification project involves user-based classification (and not all do, some rely wholly on automated data classification techniques), getting the users on board ahead of the project 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.

- 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 on classification is the foundation needed for success.

Part Seven

Getting Successful With Data Classification

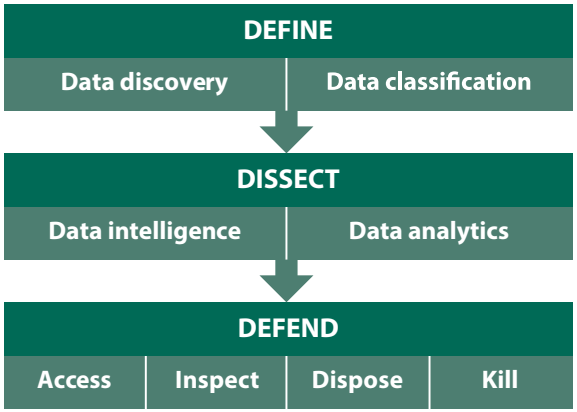
Data Protection Framework

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.



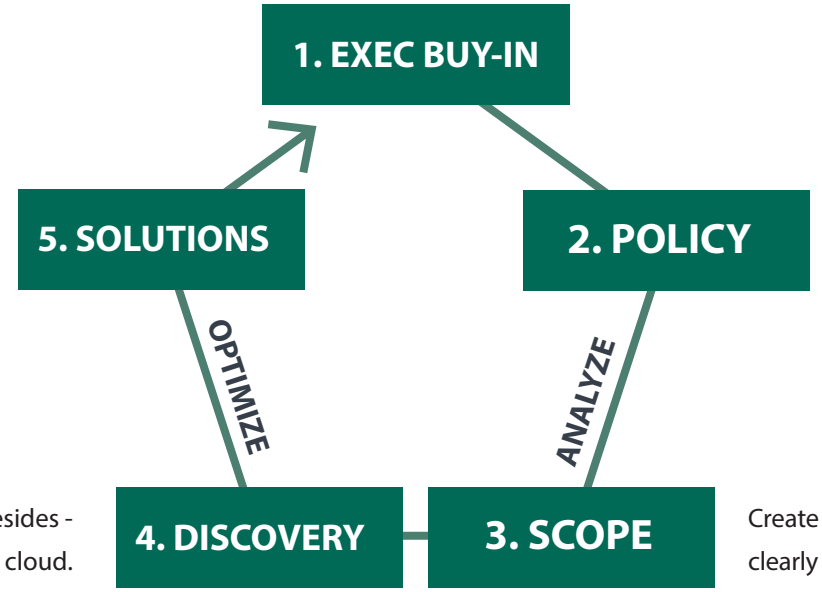
(source: The Future Of Data Security And Privacy: Growth And Competitive Differentiation, Forrester Research, Inc., October 16, 2019, Heidi Shey, Enza Iannopollo)

FORRESTER®

Data Classification Process

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

Communicate to the leadership team how classification can support increased revenue, reduced expense, and reduced risk.



Document the goals, objectives, and strategic intent behind the classification projects.

Create guardrails around your program; clearly define what is in and out of scope.

Find sensitive data - wherever it resides - including endpoint, database, and cloud.

Put in place technical and procedural controls to enforce policies.

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. Only if those three prove insufficient should you add more categories.

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!

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."

Table 1: Three Categories of Data, Classified by Risk

Enlarge Table

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.

(source: Gartner, How to Overcome Pitfalls in Data Classification Initiatives, 21 April 2020)

Part Eight

Digital Guardian Data Classification & Protection

DG 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.

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

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.

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

DG Data Classification

Digital Guardian classifies via context, content, and user based to cover the spectrum from fully automated to fully manual classification.

Digital Guardian's data classification integrates into our data protection suite. 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 Digital Guardian provides the comprehensive data protection needed to stop data theft.

EXAMPLE METHODS

CONTENT

Digital Guardian scans a database, 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.

Digital Guardian
Data Classification

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.

Digital Guardian
User Classification

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.

Digital Guardian
Data Classification

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. Digital Guardian delivers classification options that cover the spectrum from fully automated to fully manual 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.
- Manual, user-based 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 manual 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, DG proactively finds, classifies and tags files.

Digital Guardian
Data Classification

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.

Digital Guardian
Data Classification

Fully user-driven

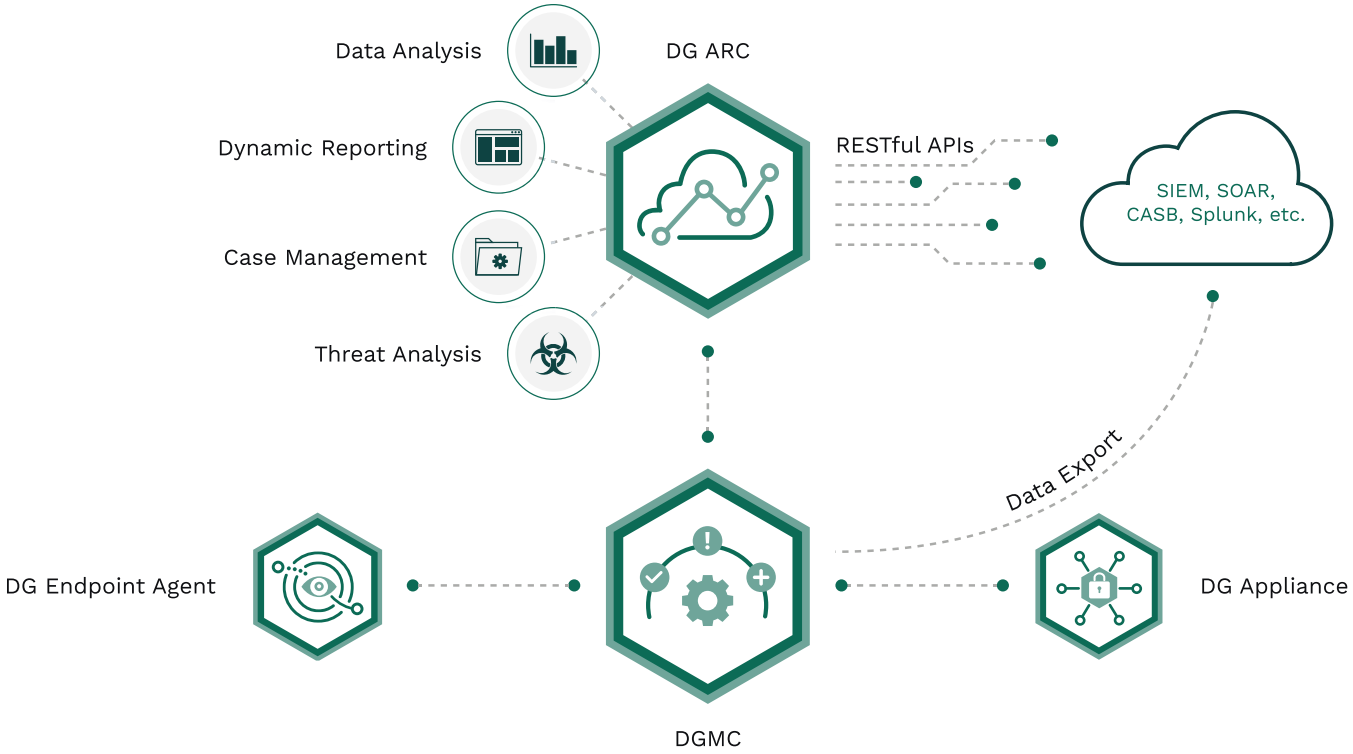
User classification relies on the data owner to apply the tag to the document at creation, or after modification.

Digital Guardian
User Classification

The Only Cloud Delivered Data Protection Platform

Data protection is at the core of our company mission. The DG Data Protection Platform detects threats and stops data exfiltration from both well-meaning and malicious insiders as well as external adversaries.

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



FREE DOWNLOAD

· Digital Guardian Platform Technical Overview

Gartner And Forrester DLP Leader

The Digital Guardian endpoint covers DLP, advanced threat protection, and endpoint detection and response (EDR) in a single agent form factor installed on desktops, laptops and servers running Windows, Linux and Mac OS X, as well as support for VDI environments...

Gartner Magic Quadrant for Enterprise Data Loss Prevention, February 2017

Digital Guardian brings together two in-demand enterprise security capabilities today: DLP and endpoint visibility and control (EVC). A strong focus on strategic partnerships augments the company's information management capabilities. It also has a popular DLP-as-a-managedservice offering that now includes local UK and EU hosting options.

Forrester WaveTM: for Data Loss Prevention Suites, Q4 2016

FORTRA

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.

fta-dg-gd-0624-r2-an