

How Is Data Classified Into Different Categories?

Roughly [88% of organizations](#) do not have confidence in their ability to detect and prevent the loss of sensitive data. This is understandable when you consider the sheer amount of data that businesses develop on a daily basis. Yet, every text message, email, image, invoice, and so much more are potential threats to a company—especially when that data is not properly classified and stored. The fact is, data that gets into the wrong hands can spell disaster for businesses.

It's not a stretch to say that sensitive data discovery and classification is mission critical for businesses operating today. The ability to properly track and assign data classification levels to data—especially unstructured data, which doesn't fit into traditional databases—can help you limit risk *and* improve security compliance.

Continue reading to learn more about how data is classified and how data discovery and classification tools can take the guesswork out of the process.

What Is Data Classification?

Data classification is the process of discovering, evaluating, and organizing data into categories by assigning tags—either in the document, with metadata, or by some combination of both. Data classification is based on factors like file type, content, context, or sensitivity level. This can serve various purposes, including:

- Improving users' ability to find and use the data they need.
- Enforcing data protection policies that ensure only authorized users have access to data.
- Mitigating risk management and meeting regulatory requirements, such as the General Data Protection Regulation (GDPR) or the Health Insurance Portability and Accountability Act (HIPAA).

When most people think of data classification, they think of rows and columns in spreadsheets. This is typically *structured data*—or highly-standardized data (often numerical in nature) that can be easily and safely stored in a database.

However, data classification is also critical to the storage and protection of *unstructured data*. Unstructured data is essentially everything else—all the information and documents that are not easily stored in a traditional database. This includes things like audio files, videos, text files, social media posts, and more. In the remainder of this article, we'll focus on how to classify the overwhelming amount of unstructured data that most businesses compile.

What Is the Best Way to Classify Data?

The best way to classify data is by implementing a [data discovery](#) and classification framework that fits your company's unique needs. Not only will this ensure that you're meeting any goals or required regulations, but it can also assist in the process of gaining buy-in from pertinent internal and external stakeholders. Here are some of the first steps to take as you begin classifying your data.

Define Your Objectives

What are your goals around discovering and classifying your data? One of the most common objectives is to ensure that you are meeting compliance regulations that exist in your industry. For example, the California Privacy Rights Act ([CPRA](#)) enhances the protections provided by the California Consumer Privacy Act ([CCPA](#)). Beginning in July 2023, companies that do business with consumers in California will be required to meet

stricter restrictions around collecting and storing of special personal information, including private consumer communications like emails or text messages.

Discover Your Data

Before moving forward, it's essential to understand what kinds of unstructured data you have and where it currently lives. Without this key process, any categories you create and implement will be based on sheer guesswork, rather than what actually exists. In this phase, you will compile all your company's unstructured data from across different environments or repositories. This might include internal data, public data, data from your customers, data from your vendors, and more.

Create Data Classification Categories

Once you've discovered your data, you can begin developing data classification tags. There are a number of different classification categories. But, what are the main types of data classification? Here are two of the most common:

Content Vs Context Based Classification

Data can either be classified as content-based or as context-based.

- **Content-based data classification** aims to sort data based on the content of the document itself, how sensitive the information is, and why types of security or authorization need to be in place.
- **Context-based data classification** aims to sort data based on various attributes, such as the application, file type, geographic location, or the creator of the file.

Sensitive Data Level

Data can be classified as having low, medium, or high sensitivity.

- **Low sensitivity data** is for public use and consumption, and it poses little to no security threat. This might include website data, annual reports, or press releases.
- **Medium sensitivity data** is for internal use. While this data could pose a potential security threat if it fell into the wrong hands, the result would not be catastrophic. This might include internal emails with no confidential or personal identifiable data (PII), or non-confidential research data.
- **High sensitivity data** includes the most confidential information. Only those who are authorized should be able to access it, and a breach could cause significant harm. What's more, high sensitivity data often has stringent controls as defined by various regulatory bodies. This might include health records, PII financial information, or Social Security numbers.

Develop an Action Plan

Finally, it's time to present your findings and recommendations to create and implement a data classification policy for your organization. In this stage, you should identify how best to organize your categories and how to use them to make business decisions moving forward. Additionally, it's essential to develop a plan to classify any new data to ensure that your data classification remains consistent over time.

How Do You Implement Data Classification?

Fortunately, you're not alone when it comes to performing all of the above steps. That's where data discovery and classification tools like DryvIQ come in handy. Our proprietary artificial intelligence (AI) models can classify content by discovering sensitive and high risk data. From there, our platform can apply metadata, document classification, or other identifying tags and labels to your unstructured data, including data entities such as:

- Document type, including resumes, W-2s, invoices, and more

- Personal identifiable information(PII), including names, addresses, dates of birth, Social Security numbers, banking information, and more.
- More than 5,000 standard government forms
- Foreign language detection
- Any custom data attributes unique to the needs of your company

With DryvIQ, you get an unstructured data solution that helps protect your company—and your data—now and in the future. Interested in learning more? [Schedule a demo](#) to see the DryvIQ platform in action.