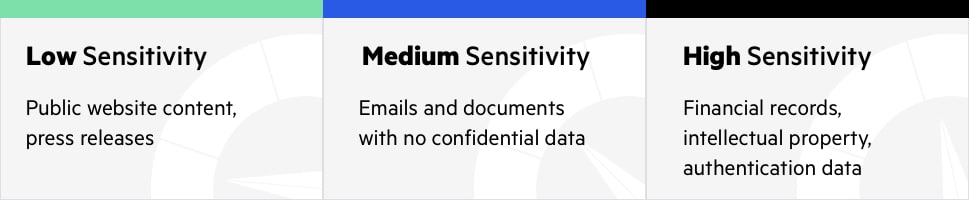
**Data Classification**

Data classification tags data according to its type, sensitivity, and value to the organization if altered, stolen, or destroyed. It helps an organization understand the value of its data, determine whether the data is at risk, and implement controls to mitigate risks. Data classification also helps an organization comply with relevant industry-specific regulatory mandates.

**Data Sensitivity Levels**

Data is classified according to its sensitivity level—high, medium, or low.

* **High sensitivity data**—if compromised or destroyed in an unauthorized transaction, would have a catastrophic impact on the organization or individuals. For example, financial records, intellectual property, authentication data.
* **Medium sensitivity data**—intended for internal use only, but if compromised or destroyed, would not have a catastrophic impact on the organization or individuals. For example, emails and documents with no confidential data.
* **Low sensitivity data**—intended for public use. For example, public website content.



**Data Sensitivity Best Practices**

Since the high, medium, and low labels are somewhat generic, a best practice is to use labels for each sensitivity level that make sense for your organization. Two widely-used models are shown below.

| **SENSITIVITY** | **MODEL 1** | **MODEL 2** |
| --- | --- | --- |
| **High** | Confidential | Restricted |
| **Medium** | Internal Use Only | Sensitive |
| **Low** | Public | Unrestricted |

**Types of Data Classification**

Data classification can be performed based on content, context, or user selections:

* **Content-based classification**—involves reviewing files and documents, and classifying them.
* **Context-based classification**—involves classifying files based on meta data like the application that created the file (for example, accounting software), the person who created the document (for example, finance staff), or the location in which files were authored or modified (for example, finance or legal department buildings).
* **User-based classification**—involves classifying files according to a manual judgement of a knowledgeable user. Individuals who work with documents can specify how sensitive they are—they can do so when they create the document, after a significant edit or review, or before the document is released.

**Data States and Data Format**

Two additional dimensions of data classifications are:

* **Data states**—data exists in one of three states—at rest, in process, or in transit. Regardless of state, data classified as confidential must remain confidential.
* **Data format**—data can be either structured or unstructured. Structured data are usually human readable and can be indexed. Examples of structured data are database objects and spreadsheets. Unstructured data are usually not human readable or indexable. Examples of unstructured data are source code, documents, and binaries. Classifying structured data is less complex and time-consuming than classifying unstructured data.

**Data Discovery**

Classifying data requires knowing the location, volume, and context of data. Most modern businesses store large volumes of data, which may be spread across multiple repositories:

* Databases deployed on-premises or in the cloud
* Big data platforms
* Collaboration systems such as Microsoft SharePoint
* Cloud storage services such as Dropbox and Google Docs
* Files such as spreadsheets, PDFs, or emails

Before you can perform data classification, you must perform accurate and comprehensive data discovery. Automated tools can help discover sensitive data at large scale. See our article on [Data Discovery](https://www.imperva.com/learn/data-security/data-discovery/) for more information.

**The Relation Between Data Classification and Compliance**

Data classification must comply with relevant regulatory and industry-specific mandates, which may require classification of different data attributes. For example, the Cloud Security Alliance (CSA) requires that data and data objects must include data type, jurisdiction of origin and domicile, context, legal constraints, sensitivity, etc. PCI DSS does not require origin or domicile tags.

**Creating Your Data Classification Policy**

A data classification policy defines who is responsible for data classification—typically by defining Program Area Designees (PAD) who are responsible for classifying data for different programs or organizational units.

The data classification policy should consider the following questions:

* Which person, organization or program created and/or owns the information?
* Which organizational unit has the most information about the content and context of the information?
* Who is responsible for the integrity and accuracy of the data?
* Where is the information stored?
* Is the information subject to any regulations or compliance standards, and what are the penalties associated with non-compliance?

Data classification can be the responsibility of the information creators, subject matter experts, or those responsible for the correctness of the data.

The policy also determines the data classification process: how often data classification should take place, for which data, which type of data classification is suitable for different types of data, and what technical means should be used to classify data. The data classification policy is part of the overall information security policy, which specifies how to protect sensitive data.

**Data Classification Examples**

Following are common examples of data that may be classified into each sensitivity level.

| **Sensitivity Level** | **Examples** |
| --- | --- |
| **High** | Credit card numbers (PCI) or other financial account numbers, customer personal data, FISMA protected information, privileged credentials for IT systems, protected health information (HIPAA), Social Security numbers, intellectual property, employee records. |
| **Medium** | Supplier contracts, IT service management information, student education records (FERPA), telecommunication systems information, internal correspondence not including confidential data. |
| **Low** | Content of public websites, press releases, marketing materials, employee directory. |