Data Classification

What is it	Where is it and what to do
Understanding your data landscape is the first step on the journey to manage and govern that data. Every day, we create, edit, and consume data in different formats, from different sources. This data can be benign and nothing more than pleasantries and chat between our employees, or it can be business sensitive, personal information, intellectual property, and more. Understanding what this data is and where it lives is critical to the success of any information governance program.	Navigate to compliance.microsoft.com Login with the credentials you have been provided for your lab tenant Click on **Data Classification** on the left hand fly out section.
The overview page in data classification provides a snapshot view of your program. On this page, you can see what types of sensitive information are most widely used and stored in your organization, across the cloud and on-premises environments. The top sensitive info types report provides a graphical view of the types of information mostly widely detected or used in the organization.	Hover over the **Top Sensitive info Types report.** Each part of the bar chart indicates a Sensitive Information Types (SIT) that has been found or detected in your environment.
Top sensitivity labels applied to the content reports give you a view of which classification labels are most widely used in the organization. This helps you understand the level of sensitivity of the data held within the organization. The classification labels will vary by organization based on your classification taxonomy and internal policies. You can configure the solution and deploy your sensitivity labels across the enterprise.	
Top retention labels applied to content provides a view of which retention policies are most widely used. Depending on legal/regulator mandates and internal policy, you might have different retention policies and requirements for your data. Financial data might have to be retained for 10 years, while personal data is only retained for 2. Most organizations will have a default retention policy that applies to general data created and shared, layered on top with additional policies to handle special data as per legal/regulatory need.	

What is it	Where is it and what to do
In the current landscape, holding on to stale data for longer than necessary does not only present a data storage cost, but also presents a risk to organizations from a data management and potential mishandling perspective.	
The dashboard provides additional reports that showcase where most data resides providing a view of cloud and on-premises breakdown. As well as activity reports from detected user actions.	
So, how does the system recognize which data to flag? Sensitive information types functionality Identifies sensitive data by using built-in or custom regular expressions or a function. Corroborative evidence includes keywords, confidence levels, and proximity. Microsoft provides 250+ predefined sensitive information types out of the box that are designed to identify most commonly used data formats related to financial, personal and medical information spanning geographies and regions around the world. However, our organizations create and hold custom information and data that is specific to our businesses. For this purpose, you are able to create your own sensitive information types. Sensitive information types can be defined by using regular expressions, keywords or functions. They can also be associated with supporting elements to increase true positive results.	Click on the **Sensitive info types** tab along the top of the dashboard. 1. Click on **+ Create sensitive info type** 2. Provide the name and description **Customer ID & (entera number of your choice from 10-100)** to make unique to your setup. 3. Click **Next** 4. Click **+ Create pattern** 5. Under Primary element click **+ Add primary element** and select **Regular expression** 6. On the new flyout pane, enter **Customer ID** in the ID field 7. Input the following as the regular expression **[0-9]{6}[a-zA-Z]** 8. Click **Done** 9. Click **Create** 10. Click **Next** 11. Click **Next** 12. Click **Create**
However, not all organizational data can be expressed with a formula, keyword(s), or regular expression. Some files and documents might be sensitive by the nature of what they are and not only the content. For example, resumes, source code, contractual agreements, minutes of meetings, and other files are sensitive by their nature regardless of the content. Trainable classifiers are more about using a classifier to identify an item based on what the item is, not by elements that are in the item (pattern matching). A classifier learns how to identify a type of content by	Click on **Trainable Classifiers** on the tab set at the top of the page. Click **Cancel** on the pop-up message **Get started with trainable classifiers** You can drill into any of the trainable classifiers to see - Details and description - Matched items detected - Provide feedback on detected items (positive match, false match)

What is it	Where is it and what to do
looking at hundreds of examples of the content you're interested in classifying. For example, this is looking for 6 numeric digits followed by 1 alpha (eg 123456A)	Note that some of the classifiers are available in multiple languages as well.
Some sensitive organizational data is dynamic and may continuously grow or change. In some cases, the information must match an exact data value. This is where Exact Data Match based classification comes in. Exact Data Match enables you to create custom sensitive information types designed to: - Be dynamically and easily refreshed - Be more scalable - Work with structured sensitive data	
Getting a view of which types of information have been detected in the organization and where is critical to understanding your data landscape. Content Explorer enables you to natively view items that were summarized on the overview page. Access to it is strictly controlled with Role-Based Access Control, and only persons with specific roles can view the content and the lists. Content explorer provides a view of which SITs (Sensitive Information Types) are present in your cloud environment, which sensitivity labels have been applied, and which retention labels are also in use.	1. Click on **Content Explorer.** 2. On the left-hand panel, select **Credit Card Number** from the list. 3. In the pane on the right, double click **SharePoint.** 4. Drill into the first folder 5. Double click on one of the documents in the list
Content Explorer provides a good view of what your data landscape looks like. To see changes and interactions with the data, we can use the activity explorer. Activity Explorer provides a historical view of activities on your labeled content. The data in here is collected from the M365 unified audit logs, transformed and provided in the activity explorer UI.	Click on **Activity Explorer** in the tab set. You can use the **built-in filters** to filter the list of activities, change the time frame for the activity and use any of the other available filters to slice and dice the data in the report