Part 1 – Sensitive Information Types (SITs)

Contents

Disclaimer

This document is not meant to replace any official documentation, including those found at docs.microsoft.com. Those documents are continually updated and maintained by Microsoft Corporation. If there is a discrepancy between this document and what you find in the Compliance User Interface (UI) or inside of a reference in docs.microsoft.com, you should always defer to that official documentation and contact your Microsoft Account team as needed. Links to the docs.microsoft.com data will be referenced both in the document steps as well as in the appendix.

All of the following steps should be done with test data, and where possible, testing should be performed in a test environment. Testing should never be performed against production data.

Target Audience

The Sensitive Information Type (SIT) section of this blog series is aimed at Compliance officers who need to identify any PII and PHI data in their environment.

Document Scope

This document is meant to guide an administrator who is "net new" to Microsoft E5 Compliance through:

- Creation of a Sensitive Information Type (SIT).
- Modification of a Sensitive Information Type (SIT).
- Testing of your Sensitive Information Type (SIT).

Out-of-Scope

This document does not cover any other aspect of Microsoft E5 Compliance, including:

- Exact Data Matches
- Data Protection Loss (DLP) for Exchange, OneDrive, Devices
- Microsoft Cloud App Security (MCAS)
- Records Management (retention and disposal)
- Information Protection
- Advanced eDiscovery

It is presumed that you have a pre-existing of understanding of what Microsoft E5 Compliance does and how to navigate the User Interface (UI).

Overview of Document

1. Use Case
2. Definitions
3. Notes
4. Pre-requisites
5. Create a new Sensitive Information Type (SIT)
6. Modify an existing Sensitive Information Type (SIT)
7. Test a Sensitive Information Type (SIT)
8. Appendix and Links

Use Case

Sensitive Information Types (SIT) are used to flag data for Compliance based upon the content of the file or email, regardless of their location. So the Use Case here is to create a SIT that does not exist out-of-the-box OR to modifying an existing SIT that is lacking a keyword or pattern that needs to have a compliance policy applied to it.

Definitions

1. Data Classification
    a. The core of the Compliance tool is the Microsoft Information Protection (MIP) engine. This engine allows for indexing of existing data and then track any changes made to that data via the Compliance tool set (example – information label that data with sensitivity and governance labels).

2. Trainable classifiers
    a. Trainable classifiers leverage Machine Learning (ML) to improve upon pre-built or "net new" keyword dictionaries. Here are 3 use cases:
        i. Harassment and bad behavior – what is determined to be bad behavior or harassment changes over time. ML allows for the system to learn this as more data is put into the Tenant.
        ii. Vulgarity – not all vulgarity is known. Often these are slang terms that no one organization knows on day one.
        iii. Applications – Most organizations leverage standardized applications for admission, release forms, etc. Only certain fields change (example – name, address, data of birth, etc).

3. Sensitive Information Types (SITs)
    a. A SIT is anything that might be unique that you want to track, label, block or set a retention label on. For example – Credit Card information, Passport information, Social Security Numbers, Medical Record Numbers, Patient IDs, etc.
    b. You can create new SITs and you can modify SITs that you create.
    c. You cannot modify an MS out-of-the-box SIT.

4. Exact Data Matches (EDMs)
    a. EDMs are used when you want to group large amounts of data specific to your organization. As the name infers, these are unique to the organization. For example, instead of any SSN or even any 9 digit number, it would be a unique and specific SSN. Here are a few examples of how EDMs this would be used.
        i. Example #1 – PII specific to customers
            1. Customer information such as Employee numbers, Social Security Numbers, Credit Card numbers, addresses, Dates of Birth, etc.
        ii. Example #2 – PHI specific to patients

1. Patient information such as Medical Record Numbers, Patient IDs, Social Security Numbers, Credit Card numbers, addresses, Dates of Birth, etc.

5. Content Search
    a. This allows for a compliance officer to see where the data resides in an organization's environment. This does not access the data directly. It is searching the indexes created by EXO, SPO and OneDrive by default. Also, using this search provides for the compliance officer to see where the data is, review the context, if necessary, etc.
        i. Note – It does not allow for eDiscovery search, hold, and export. Nor does it allow for them to apply labels or encryption. It also does not allow for them to apply Data Loss Policies. Neither does it allow them to apply retention and disposal policies. That is all done in other parts of the Compliance UI.

6. Activity Explorer
    a. This allows the compliance officer or IT team to view which activities have been run in the Compliance tool and then drill down into what was done, when, and by whom. It should be viewed as an auditing tool.

Notes

- Replication times for a Compliance changes to take affect
    - DLP policies will take approximately 15 minutes to take affect
    - Other Compliances items could take 24-48 hours for other changes to take affect

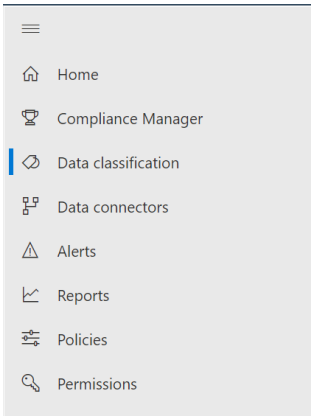
Requirements

Your test account will need the following rights to run the activities in this blog series. See the link in the Appendix below for the link for eDiscovery rights.

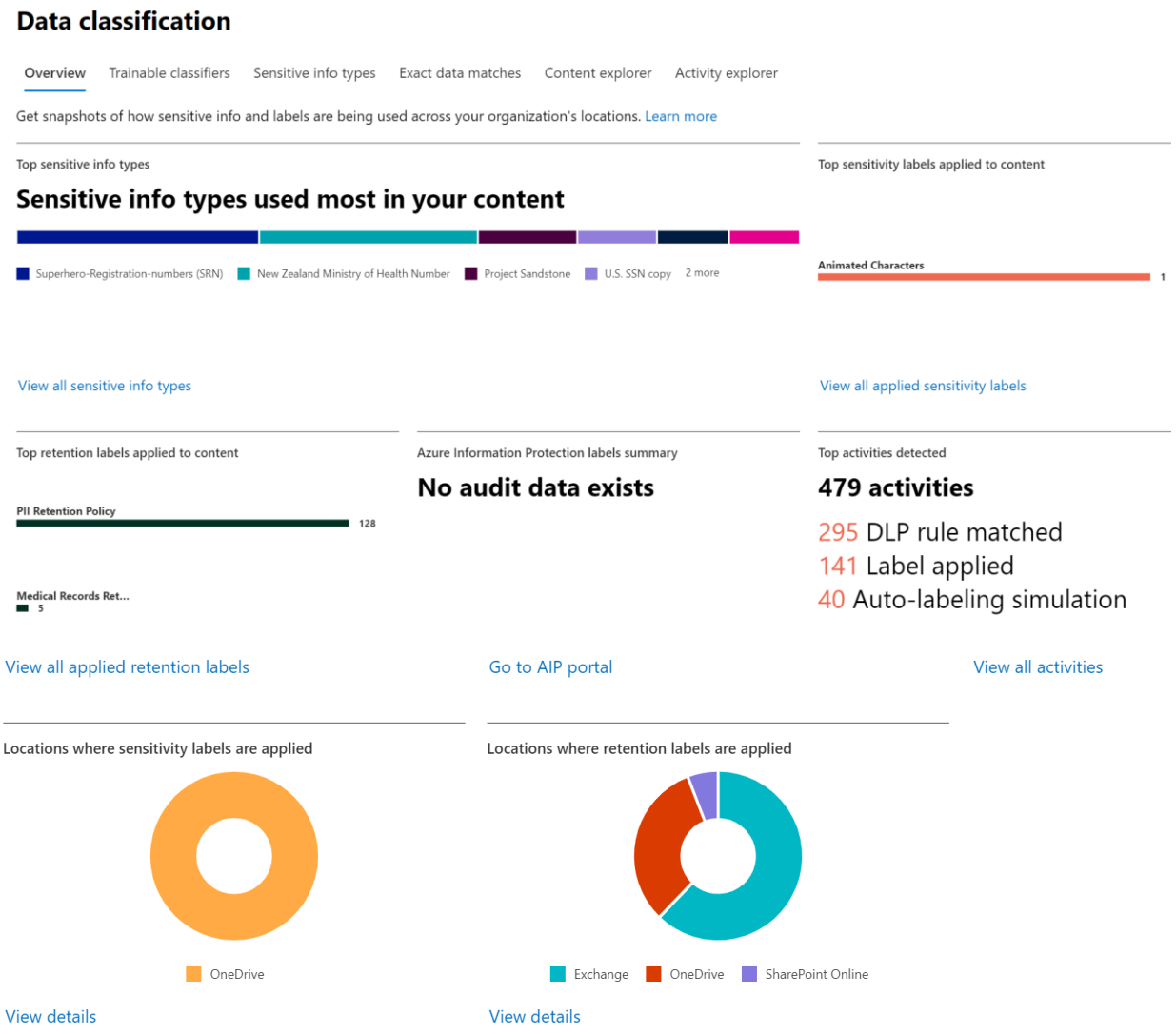
- Compliance Administrator
- eDiscovery Administrator
- eDiscovery Manager

Data Classification Overview

1. In the left-hand navigation pane, select **Data Classification**



2. In the right-hand pane, you will see several options. The first is **Overview**. This is a dashboard view of your Sensitive Information Types (SITs) and where they are located in your environment.

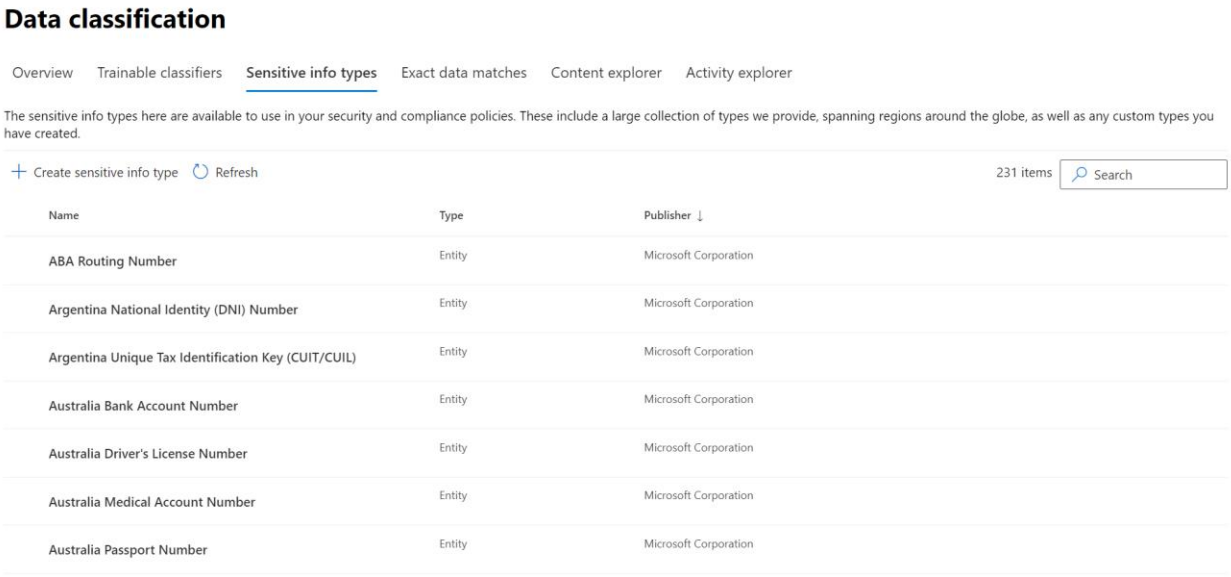


3. You will see links in each sub-pane in this dashboard. These will allow you do drill down into the information provided, or you can click on the links across the top. We will not be looking at all the tabs, only the **Sensitive information types** tab.

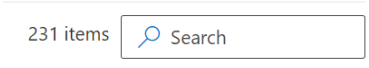
Sensitive Information Types (SITs)

Before we create or modify a SIT, let us look at the SIT pane and an existing SIT.

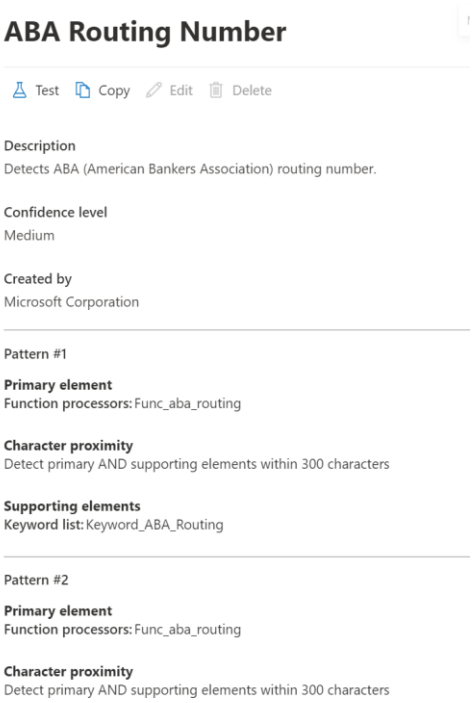
1. Click on the **Sensitive Info Types** (SITs).



2. In the top right-hand corner, you will find a search field where you can search all of the SITs in your Tenant.



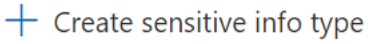
3. If you click on one of these SITs, a window will pop-up on the right showing the details of the SIT. You cannot modify an out-of-the-box SIT, but you can copy one and modify that copy. More on that later on. You can also run a test on a file to verify that the SIT is seeing your data. Again, more on that later on. Here is an example of that pop-up.



Creating a Sensitive Information Type (SIT)

First, we will create a new SIT.

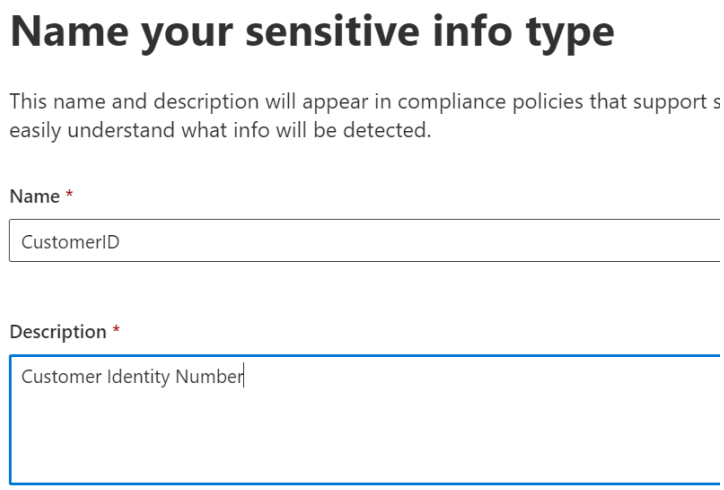
1. Click **Create info type** in the top-right of the work pane



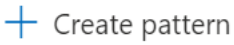
2. You will start a 4-step wizard.

3. First, give a Name and Description to your new SIT.

    a. Example – Name – CustomerID
    b. Example – Description – Customer Identity Number



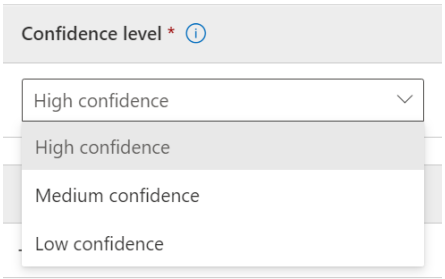
4. Next you will create a pattern to associate with your SIT. Click **Create pattern**.



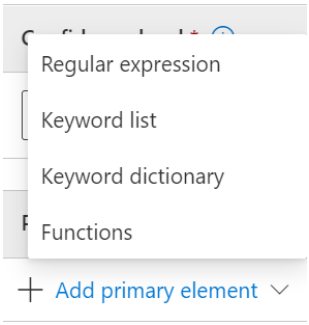
5.

6. A **New Pattern** pane will appear on the right-side of the screen.

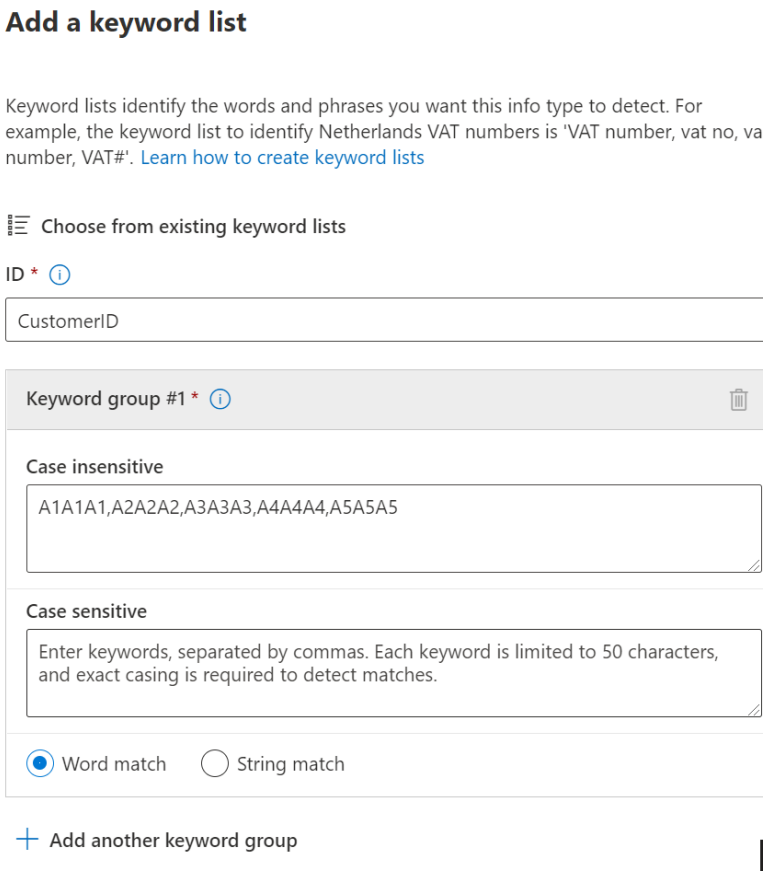
7. Choose your confidence level for the SIT you are about to create (High, Medium, or Low). I am going to choose **High Confidence** for my SIT.



8. Click **Add Primary Element**. You can choose from Regular Expressions, Keyword lists, Keyword dictionaries or Functions. I will choose a **Keyword list** as it is the simplest to create. There are **Keyword dictionaries** which are more robust keyword lists. Then there are **Regular expressions**. The most complicated element to create are **Functions**. We will not be covering those options in this document.

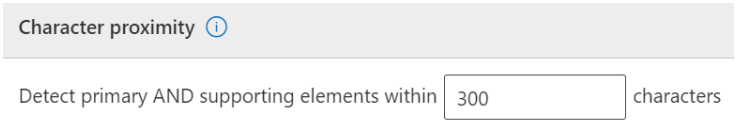


a. You can choose from an existing keyword list or create a new **ID**.
    i. Example – ID = CustomerID.
b. Under **Keyword group #1**, I am going to only fill in the **Case insensitive** field separating the words with commas. You can use any word or list of words but I will be using alphanumeric numbers. You will notice I only separated the items by commas. No spaces were used.
    i. Example – A1A1A1,A2A2A2,A3A3A3,A4A4A4,A5A5A5
c. I will be using **Word Match**, which is the default.
    i. Note – You can add more than one keyword group if desire.

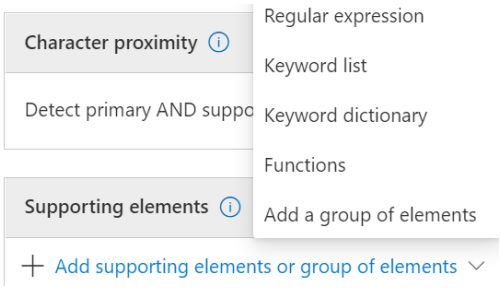


d. When you are sure of you have what you want, click **Done**.

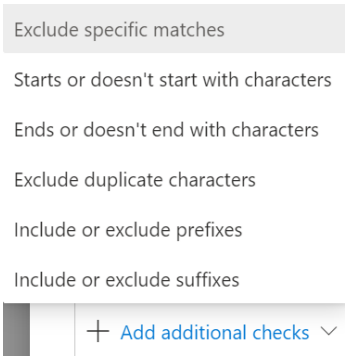
9. There is a proximity option between primary and secondary elements. We will leave this at the default of 300 characters.
    a. Note – this is 300 characters, not words.



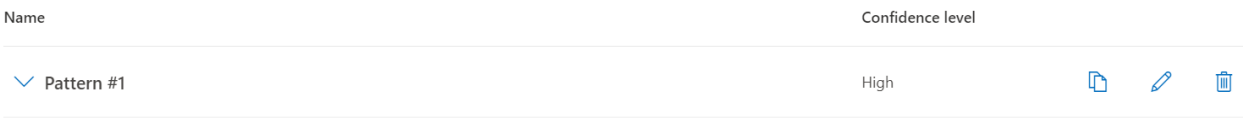
10. You can add **Supporting elements** if you like. Similar to Primary Elements, these can be Regular Expressions, Keyword lists, Keyword dictionaries or Functions. Additionally, you can add a group of elements. We will not be adding a Supporting element at this time.



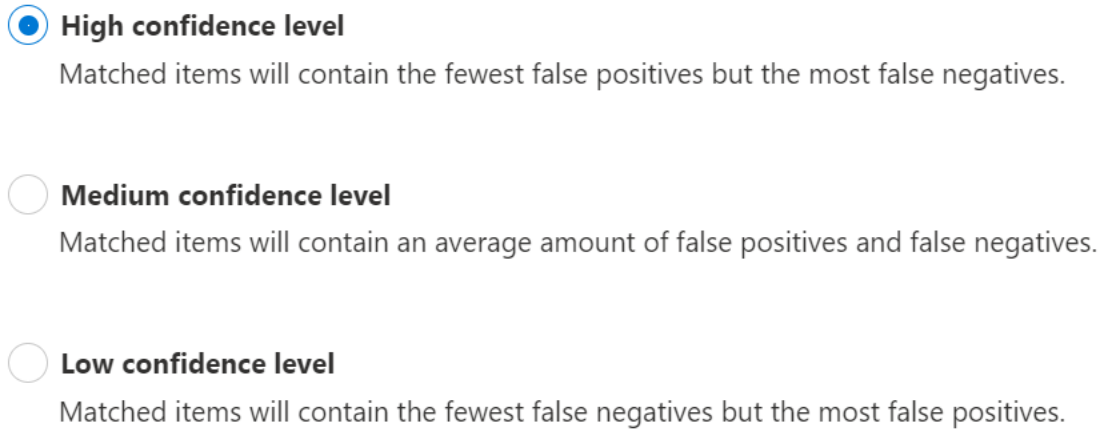
11. You can also add **Additional checks**. Below are some examples of what these can be. We will not be adding additional checks at this time



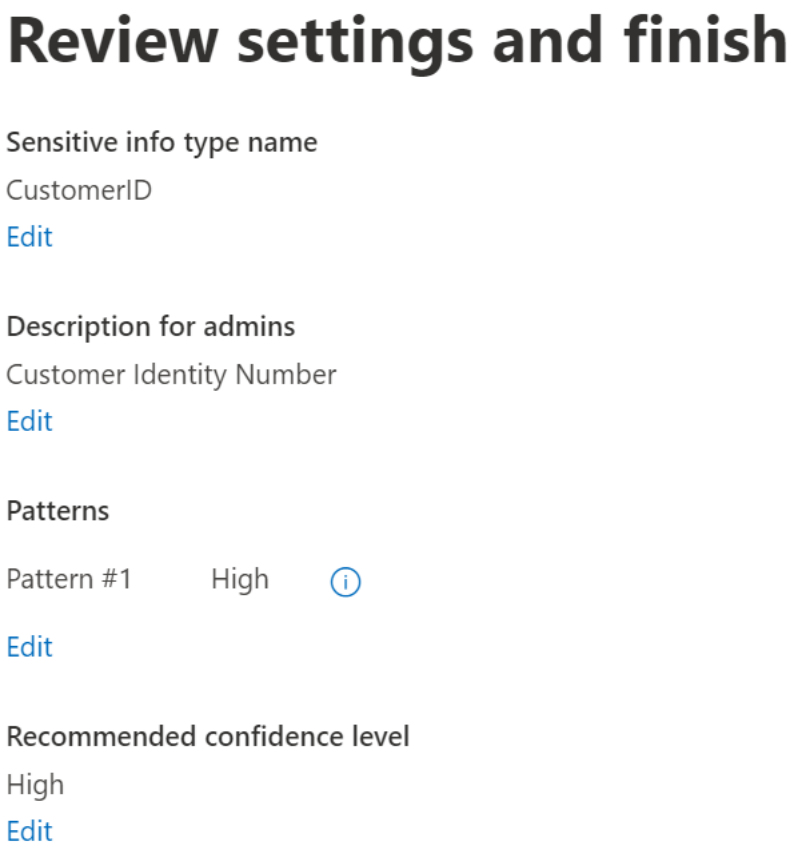
12. When you are satisfied with what you see, click **Create**.
13. You can copy, edit or delete this pattern. Copying it will create a duplicate of the pattern. You can then modify this duplicate as needed. This is useful if you are wanting to do variations of the same pattern (keywords, functions, regular expressions) as part of your SIT.



14. Click **Next** when you are ready.
15. Confirm the confidence level of the pattern when you are ready.



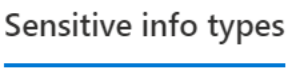
16. Click **Next** when you are ready.
17. Perform one final review of your SIT and when you are satisfied, click **Create**.



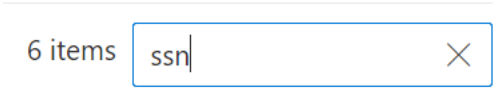
Modifying a Sensitive Information Type (SIT)

Next, we will Copy and Modify and Existing SIT. We are doing this against Social Security Numbers so we only look for the numeric pattern of those numbers without the need to have the associated keyword (ex. SSN or SocSecNum). We will use this SIT in other parts of this blog series.

1. Stay in the middle pane, click on **Sensitive information types**



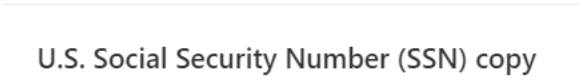
2. On the right-side of the pane enter “SSN” in the search field



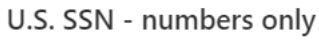
3. Select U.S. Social Security Number (SSN).

4. In the right-hand pop-up, select **Copy**.

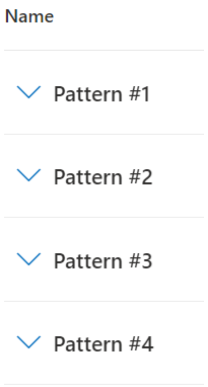
5. Select the new copy (default name will be “U.S. Social Security Number (SSN) copy”), and in the right-hand pop-up, select **Edit**.



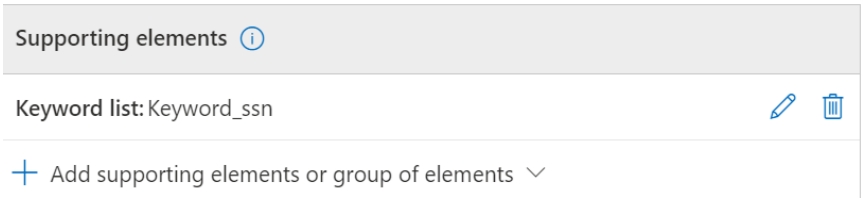
6. In the wizard that appears, click on the **Name** section, rename the copy of the U.S. Social Security Number (SSN). I have renamed mine “U.S. SSN – numbers only”.



7. In the wizard that appears, click **Next** until you arrive at the **Patterns** section of the wizard. You will find 4 patterns.



8. For each pattern, click the pencil icon to edit the pattern.
9. Find the **Supporting elements** named "Keyword_ssn" and delete it. Then click **Update**.



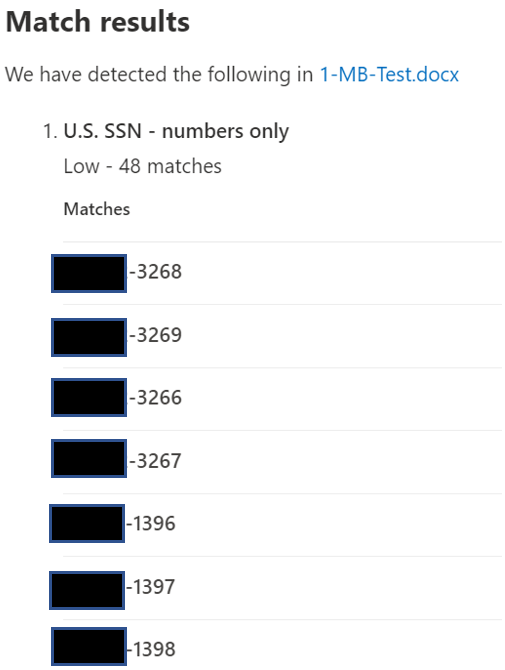
10. Repeat the step above for the other 3 patterns.
11. Once all the patterns are updated click **Next** until you get to the end and then **Save** the modified SIT.

Test a Sensitive Information Type (SIT)

1. Click on your either your new or modified SIT and click **Test** in the pop-up to the top-right.

2.
3. Click **Upload File** and browse a file with your test data.

4.
5. Click **Test**. If everything has been set up correctly, you should see something like below that I had in my "U.S. SSN – Numbers Only" test.



6. Click **Finish**.

Now that you have created, modified and test a SIT, you are ready to move onto one of the parts of this blog.

Appendix and Links

- Learn about sensitive information types - Microsoft 365 Compliance | Microsoft Docs
- Get started with custom sensitive information types - Microsoft 365 Compliance | Microsoft Docs
- Assign eDiscovery permissions in the Security & Compliance Center - Microsoft 365 Compliance | Microsoft Docs