

Microsoft Purview

Formerly the Microsoft Compliance Center

- [Sensitive Information Types](#)

Sensitive Information Types

Sensitive Information Types or SITs are pre-defined or custom patterns used to classify and identify important information that needs to be protected. Which classifiers are used are dependent on industry and necessary compliance and/or regulatory needs. Every organization will have some level of sensitive information whether it is their employees Personally Identifiable Information (PII), credit card information used in processing transactions that needs to be limited and protected for PCI, or health care information. SITs can be used in many locations in the security and compliance portals: DLP policies, sensitivity labels, retention labels, IRM, Priva, and auto-labeling policies.

There are many pre defined SITs for standardized types of data - bank account and routing numbers for different locations, social security numbers, secret keys for common cloud services, drivers license numbers, passports, blood test or other medical diagnostic codes, etc. This full list can be viewed in the Purview portal (compliance.microsoft.com) under Data Classification > Classifiers > Sensitive Information Types.

Test Sensitive Information Type

Individual documents can be uploaded to test the SIT by selecting it from the list and clicking on the test button. Only one file can be tested at a time but there should be one with a positive and one without to validate both cases.

Create Custom Sensitive Information Type

Custom SITs can be created if the predefined types don't meet your needs.

1. To do so go to the Sensitive Information Types under Data Classifiers and click Create Sensitive Information Type.
2. Enter a unique Name and a description.
3. Under Patterns Create a Pattern. The pattern must have a primary element and a confidence level, supporting elements and checks are optional.
 1. The confidence level is used to determine a match with High being certain that it is a match. High confidence reduces false positives but increases false negatives. Low confidence has more less false negatives but more false positives.
 2. The Primary Element can be a regex, keyword list, keyword dictionary, or function. Functions are groups of regexes pre-defined
 3. If you specify supporting elements, you can specify the number of characters it needs to appear within the primary element (default is 300 characters). For example, you could look for your order number and a customer name close to each other to increase confidence for matching the order number.

Copy Sensitive Information Type



While you cannot edit the default SITs, you can copy and then modify the copy.

1. Go to Compliance Portal > Data Classification > Classifiers > Sensitive Information Types
2. Select the SIT you want to copy
3. Click Copy and specify new name and description.
4. Edit existing patterns as needed.

References:

[Learn about sensitive information types | Microsoft Learn](#)

[Create custom sensitive information types | Microsoft Learn](#)