

Discover and Classify Sensitive Data in Image-Based Documents

Introduction

Sensitive personal information such as Social Security, passport, driver's license, and credit card numbers can be hidden in images and image-based documents such as PDFs, TIFFs, JPEGs, etc. Such information can escape discovery by traditional search tools like RegEx, resulting in increased exposure to loss via an intrusion or disposal of media that has not been sanitized.

All or nearly all data protection laws and regulations brought into force over the past five years require organizations to effectively identify and protect personal information within their environment.

The California Privacy Rights Act of 2020 (CPRA) cites multiple categories of personal information as "sensitive," meaning that such information cannot be used, shared, or sold without the consumer's permission. Penalties for violating the statute can be as high as \$2,500 per record, or \$7,500 per record when the personal information of a child is involved.

As a consequence, discovery and protection of personal information captured in images and image-based documents has become paramount.

Requirements

This process requires the following prerequisites in which to make this process possible. These include:

- Current versions of the SDM product(s).
- All Files "SDD".
- "Workflow" rule.
- Search "Policy".

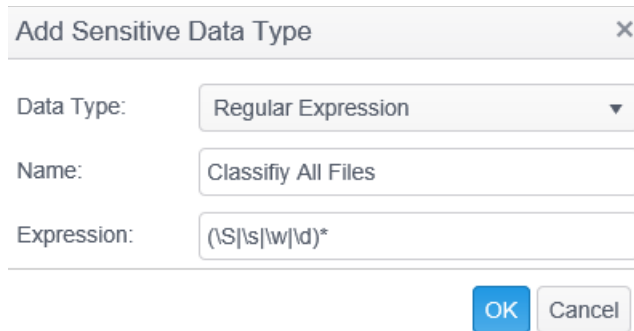
Important Notes

Note that proceeding with the following solution may override the true content-based classifications for any file in these respective locations. This Cool Solution applies to Sensitive Data Manager (SDM).

Process

STEP 1 (SDD)

1. Import or create a RegEx to capture all files in a location using the expression.
2. From within the Spirion SDM Console Admin > Sensitive Data Types > Add > Select Data Type = Regular Expression > Name = Classify All Files > Expression = `(\S\s|\w\d)*`



Add Sensitive Data Type

Data Type: Regular Expression

Name: Classify All Files

Expression: `(\S\s|\w\d)*`

OK Cancel

STEP 2 (Workflow)

1. Import or create a new workflow from within the Spirion SDM Console Admin > Classification > Add > Name = XXXX.
2. Select optionally a color, icon, or weight.
 - a. Click OK.
3. Highlight the new Classification, from within the Spirion ribbon Select Rule > Add.
 - a. Workflow Rule
 - i. Provide a workflow rule name and other options
 - b. Definition
 - i. Select from the dropdown options the following, see the image below.
 1. "Location", "Contains", Path to folder which all files from within should be classified as XXXX.

Workflow Rule	Definition	Endpoints	Actions	Assignments	Notifications
<p>The workflow rule definition specifies when the associated actions and assignments will be applied and notification whether the rule will apply to individual locations or to all of the locations found during a search. Next, specify when a single result matches all of the conditions that are joined by that AND or when each condition is satisfied or across all locations in the search, depending on the scope. Finally, specify the criteria that will trigger the rule by operation and then specifying the value. To add additional criteria, use the plus symbol to the right of the current operator (And/Or) above the desired group, as appropriate.</p> <p>Scope: <input checked="" type="radio"/> Location: Apply actions and assignments to each location that meets this rule. Send a notification <input type="radio"/> Search: Apply actions and assignments to all of the locations from a search that meets this rule.</p> <p>Type: Meet conditions linked with an AND when: <input checked="" type="radio"/> A single result matches all of the conditions <input type="radio"/> A group of results match all of the conditions</p> <p>Totals: Calculate totals across only those results that match all of the conditions.</p> <p>Definition:</p> <p>x <input type="text" value="Location"/> <input type="text" value="Contains"/> <input type="text" value="C:\Location\PoC_Test_Data"/> +</p>					

c. Endpoints

- i. Select the appropriate endpoint from where the local files are being searched.

d. Actions

- i. The Classification tag will be automatically selected, if not from the drop down for "Classify results as:" select the XXXX classification name from step 3. a. i.
- ii. Select from the "Execute classification rules:" "Directly on the endpoint".

e. Click Finish

- i. Bottom right area of the page to save the new rule.

STEP 3 (Search Policies)

1. Import or create a new workflow from within the Spirion SDM Console Admin > Policies.
2. Create a new Policy by clicking Policy > Create.
 - a. On the Policy Tab provide a name and optionally a description.
 - i. For Policy Type select Scheduled Task.
 - ii. On the Endpoints Tab select the same endpoint chosen for the workflow created previously.
 - iii. On the Data Types Tab deselect all.
 - iv. On the Location Tab deselect all.
 - b. Click Finish
3. Select the policy name just created > expand the tree view > expand Search Locations and Select Custom Folders.
 - a. Click Add from the Ribbon
 - b. In the new Folder Location field place the same Share name as done in workflow, for example (C:\Location\PoC_Test_Data).

- c. Select "Include in Search" to the right of the folder patch for the Scope.
 - d. Click the green check mark to the left of the folder path to save the changes.
4. Select the policy name just created > expand the tree view > Select Sensitive data Types.
 - a. From within the resulting list select "Classify All Files" as created in this document "Step 1, 2."

Process

STEP 1 (SDD)

1. Import or create a RegEx to capture all files in a location using the expression.
2. From within the Spirion SDM Console Admin > Sensitive Data Types > Add > Select Data Type = Regular Expression > Name = Classify All Files > Expression = (\\S|s|\\w|\\d)*

Add Sensitive Data Type ×

- b.
5. Select the policy name just created > expand the tree view > Select "Scheduled Tasks" or "Search > Initiate Search" to search and classify all files from within the target folder location.

Outcomes

As result of the following detailed procedural steps all files from within a folder will be classified as XXXX so that data can be tracked (tagged) back to the source location from which it originated.