# Automating Data Security:
# Identifying Sensitive Data with Microsoft Purview



Dutch Data Security & Compliance Community
Friday 16 May 2025 @ Dutch Technology eXperts

# Agenda

- Introduction
- Classifiers
  - Built-in and Custom Sensitive Information Types
  - Fingerprint Sensitive Information Types
  - Built-in and Custom Trainable Classifiers
  - Exact Data Match Classifiers
- Takeaways and Q&A

# About me

- CTO @ Cloud Life
- Microsoft Security MVP
- Blogger and Speaker
- Huge fantasy geek



*Myron Helgering*

in/myronhelgering    myronhelgering.com

# Why is classifying your sensitive data important?

# Which data is sensitive or important enough?

- Intellectual property

- Medical information

- Personal information

- Pricing lists

- Customer information

- Business processes

- Financial data

- Or any other type of data

# Who will classify data?

## People

## Automatic

# What can we do with classifiers?

- Classify and protect your data with Information Protection

- Prevent data loss with Data Loss Prevention

- Govern your data with Data Lifecycle Management
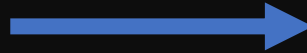
# Meet Minerva McGonagall

- Professor at Hogwarts School

- Head of Gryffindor House

- Chief Information Security Officer (CISO)

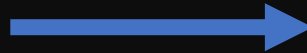Her goal is to identify, classify and protect sensitive school data.

# What does she need to start?

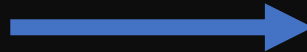- Create classifiers and policies in Microsoft Purview ➞
  - M365 E3 (or Business Prem.) license
  - Compliance administrator role

- Access to content explorer
- Apply labels automatically based on classifiers ➞
  - M365 E5 or E5 add-on license

- List and view content in content explorer ➞
  - Content Explorer Content viewer

# Built-in Sensitive Information Types

311 Built-in SITs to classify specific sensitive information;

- Credit card

- National Insurance number (UK)

- SSN number (USA)

- Passport number

- Physical address

- Ip-address

- Driver's license

- Password

## Classifiers

Trainable classifiers    **Sensitive info types**    EDM classifiers

The sensitive info types here are available to use in your security and compliance policies. These include a large collection of types we

+ Create sensitive info type    + Create Fingerprint based SIT    ↻ Refresh

| Name ↑ | | Type | Publisher |
|---|---|---|---|
| ☐ ABA Routing Number | ↗ | Entity | Microsoft Corporation |
| ☐ ASP.NET Machine Key | | Credential | Microsoft Corporation |
| ☐ All Credential Types | | BundledCredential | Microsoft Corporation |
| ☐ All Full Names | | BundledEntity | Microsoft Corporation |
| ☐ All Medical Terms And Conditions | | BundledEntity | Microsoft Corporation |
| ☐ All Physical Addresses | | BundledEntity | Microsoft Corporation |
| ☐ Amazon S3 Client Secret Access Key | | Credential | Microsoft Corporation |
| ☐ Argentina National Identity (DNI) Number | ↗ | Entity | Microsoft Corporation |
| ☐ Argentina Unique Tax Identification Key (CUIT/CUIL) | ↗ | Entity | Microsoft Corporation |
| ☐ Australia Bank Account Number | ↗ | Entity | Microsoft Corporation |
| ☐ Australia Driver's License Number | ↗ | Entity | Microsoft Corporation |
| ☐ Australia Medical Account Number | ↗ | Entity | Microsoft Corporation |

# Demo: Sensitive Information Types

I want to know how many and where (student) passport numbers are being stored

Microsoft Purview

New Microsoft Purview portal

Home

Home

Compliance Manager

Data classification

⚡ What's new ? ＋ Add cards

Overview

Classifiers

Content explorer

Activity explorer

Active alerts

Data connectors

AI hub (preview)

0 active alert

Alerts

| Alert name | Severity | Last activity |
|---|---|---|

Policies

Roles & scopes

Trials

Solutions

Catalog

# Custom Sensitive Information Types

**StudentID numbers**

- Number format is 1234-5678

- Wordlist includes StudentID

Doc 1: Harry has ID number 389-9468

Doc 2: Cho has StudentID number 0397-3845

Learn the basics Regular Expressions https://regexone.com/
Test and create Regular Expressions https://regexr.com/

# Demo: Custom Sensitive Information Types

I want to be able to classify documents that contain a StudentID

Microsoft Purview

New Microsoft Purview portal    MH

Home

Compliance Manager

Data classification
  Overview
  Classifiers
  Content explorer
  Activity explorer

Data connectors

AI hub (preview)

Alerts

Policies

Roles & scopes

Trials

Solutions

Catalog

App governance

Audit

Content search

# Home

What's new ?    + Add cards

Active alerts

## 0 active alert

| Alert name | Severity | Last activity |
| --- | --- | --- |

https://compliance.microsoft.com/dataclassificationclassifiers?viewid=sensitiveinfotypes&tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

Microsoft Purview

New Microsoft Purview portal

MH

Home

Compliance Manager

Data classification

Overview

Classifiers

Content explorer

Activity explorer

Data connectors

AI hub (preview)

Alerts

Policies

Roles & scopes

Trials

**Solutions**

Catalog

App governance

Audit

Content search

# Classifiers

Trainable classifiers       **Sensitive info types**       EDM classifiers

The sensitive info types here are available to use in your security and compliance policies. These include a large collection of types we provide, spanning regions around the globe, as well as any custom types you have created.

➕ Create sensitive info type    ➕ Create Fingerprint based SIT    🔄 Refresh

331 items

Search

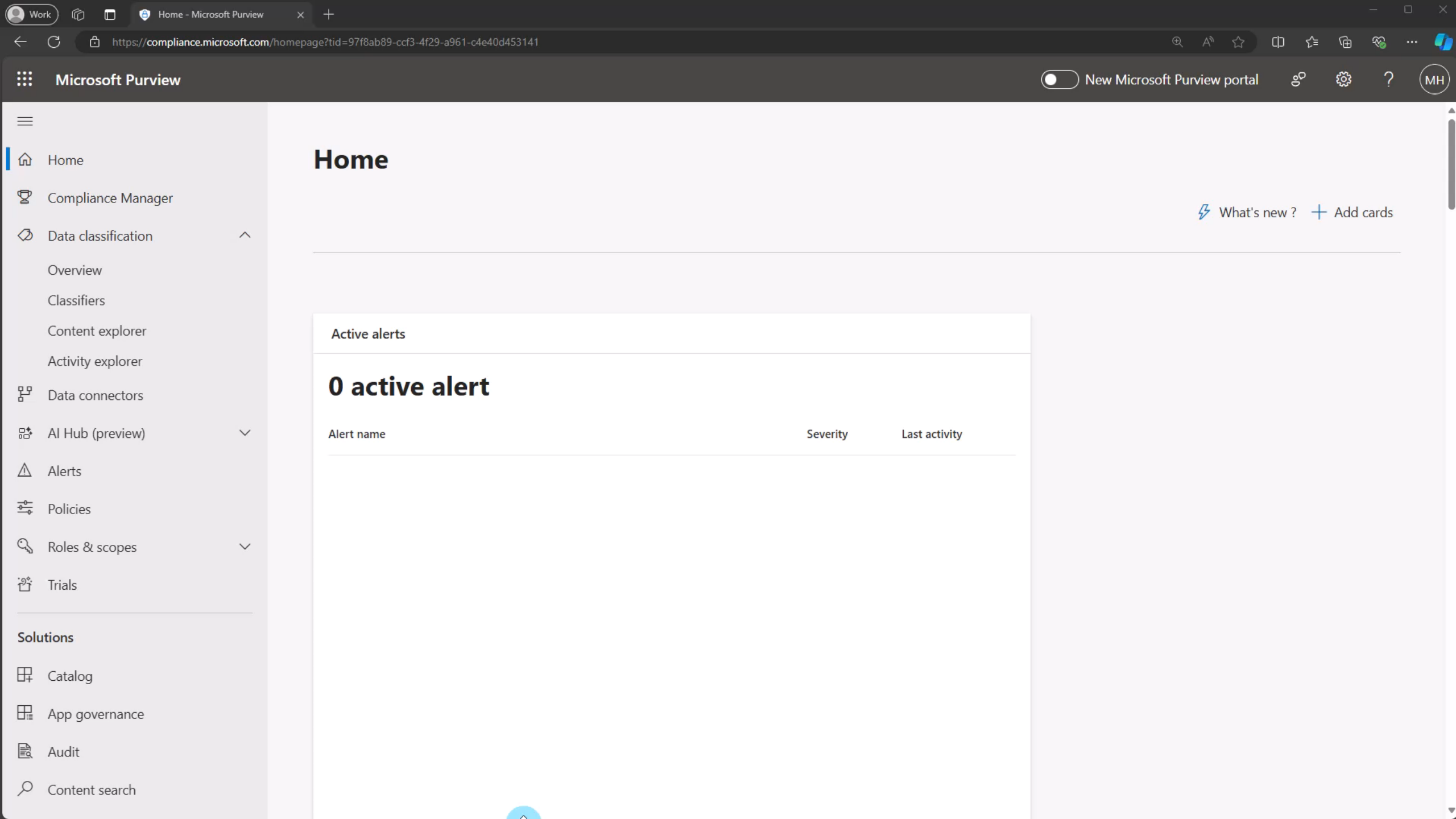| Name ↑ ⌄ | | Type ⌄ | Publisher ⌄ |
|---|---|---|---|
| ☐ ABA Routing Number | ↗ | Entity | Microsoft Corporation |
| ☐ All Credential Types | | BundledCredential | Microsoft Corporation |
| ☐ All Full Names | | BundledEntity | Microsoft Corporation |
| ☐ All Medical Terms And Conditions | | BundledEntity | Microsoft Corporation |
| ☐ All Physical Addresses | | BundledEntity | Microsoft Corporation |
| ☐ Amazon S3 Client Secret Access Key | | Credential | Microsoft Corporation |
| ☐ Argentina National Identity (DNI) Number | ↗ | Entity | Microsoft Corporation |
| ☐ Argentina Unique Tax Identification Key (CUIT/CUIL) | ↗ | Entity | Microsoft Corporation |
| ☐ ASP.NET Machine Key | | Credential | Microsoft Corporation |
| ☐ Australia Bank Account Number | ↗ | Entity | Microsoft Corporation |
| ☐ Australia Driver's License Number | ↗ | Entity | Microsoft Corporation |

# Fingerprint Sensitive Information Types
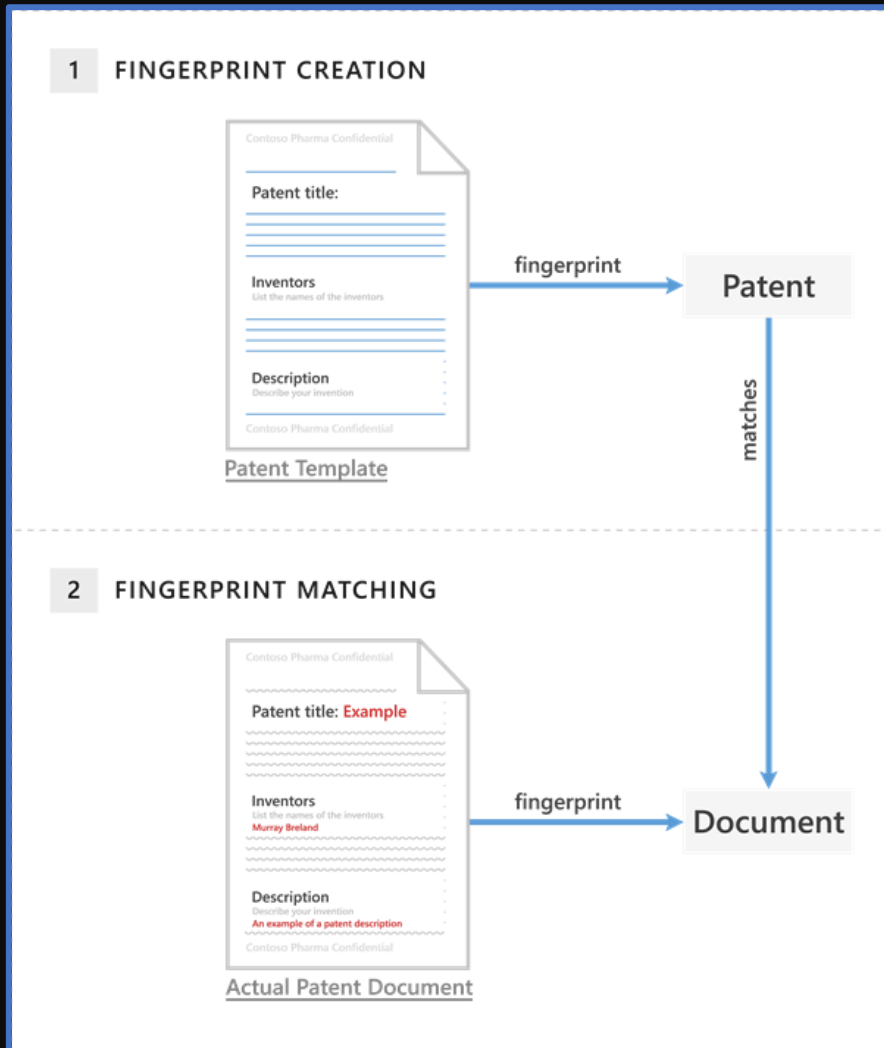


Can be used with many types of template documents, such as;

- Employee information forms

- Customer information forms

- Medical records

- Contracts

- Invoices

- Non-Disclosure Agreements (NDA)

- Or any other custom template file

Microsoft Purview

New Microsoft Purview portal

Home

Home

Compliance Manager

Data classification

Overview

Classifiers

Content explorer

Activity explorer

Data connectors

AI Hub (preview)

Alerts

Policies

Roles & scopes

Trials

What's new ?    Add cards

Active alerts

0 active alert

| Alert name | Severity | Last activity |
| --- | --- | --- |

Solutions

Catalog

App governance

Audit

Content search

# Built-in Trainable Classifiers

- Trainable classifiers identify a "type" of document
- 143 pre-trained classifiers are ready-to-use
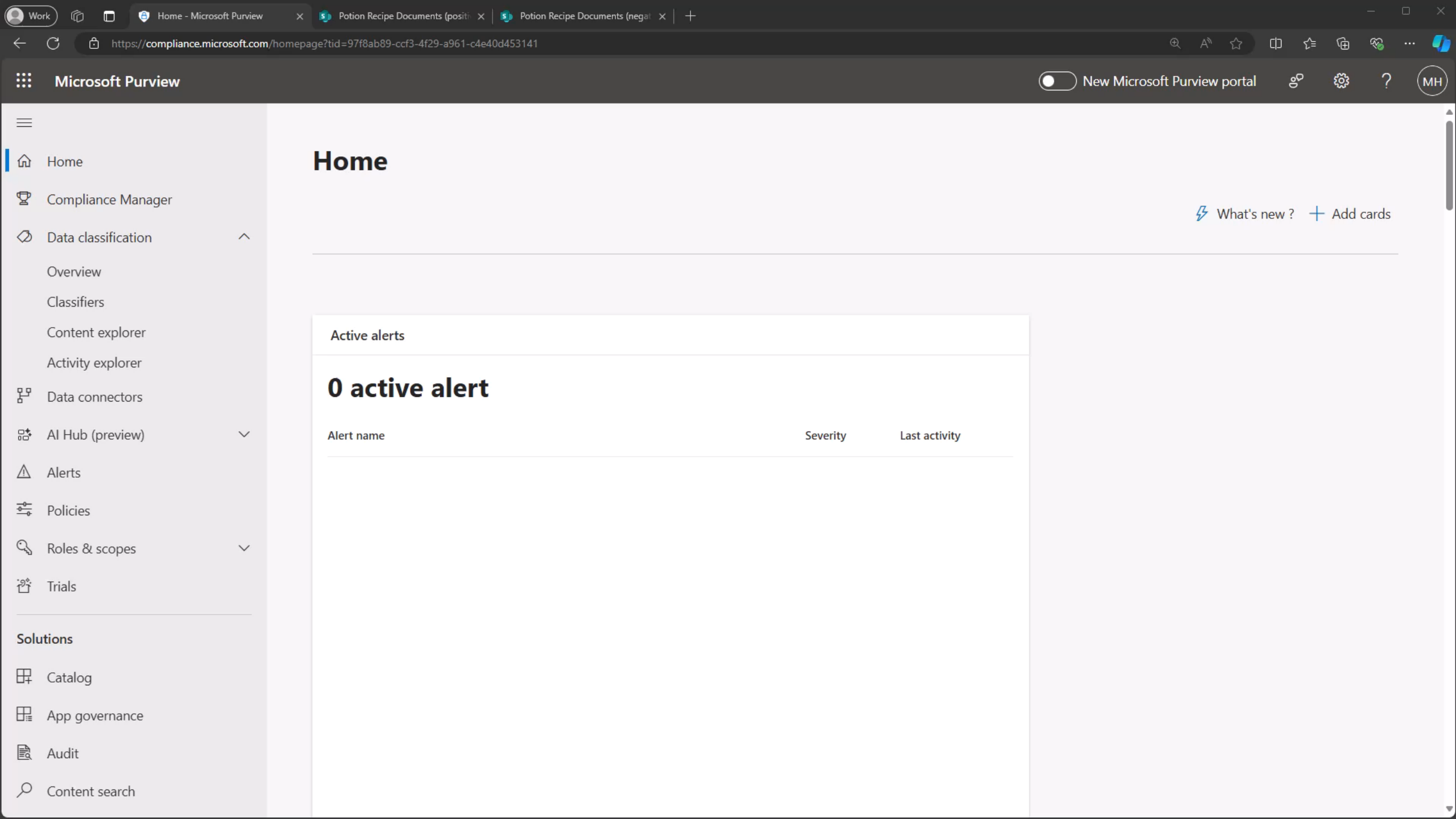- 14 different languages

# Custom Trainable Classifiers

1. 50-500 positive and negative samples

2. Microsoft trains the classifier

3. Review false negatives

4. Retrain for higher accuracy

5. Publish, test and use the classifier

# Demo: Custom Trainable Classifiers

I want to be able to identify and classify potion recipes

Microsoft Purview

New Microsoft Purview portal

MH

Home

**Home**

Compliance Manager

Data classification

Overview

Classifiers

Content explorer

Activity explorer

Data connectors

AI Hub (preview)

Alerts

Policies

Roles & scopes

Trials

What's new ?    + Add cards

Active alerts

# 0 active alert

| Alert name | Severity | Last activity |
| --- | --- | --- |

Solutions

Catalog

App governance

Audit

Content search

# Waiting….. (maximum of 48 hours)

https://compliance.microsoft.com/dataclassificationclassifiers?viewid=classifiers&tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

Microsoft Purview

New Microsoft Purview portal

MH

Home

Compliance Manager

Data classification

    Overview

    Classifiers

    Content explorer

    Activity explorer

Data connectors

AI Hub (preview)

Alerts

Policies

Roles & scopes

Trials

Solutions

Catalog

# Classifiers

**Trainable classifiers**     Sensitive info types     EDM classifiers

Use built-in or custom classifiers to identify specific categories of content based on existing items in your organization. Once created, classifiers can be used in several compliance solutions to detect related content and classify it, protect it, retain it, and more. Learn more

+ Create trainable classifier     ⟳ Refresh        135 items    ☰ Group ⌄
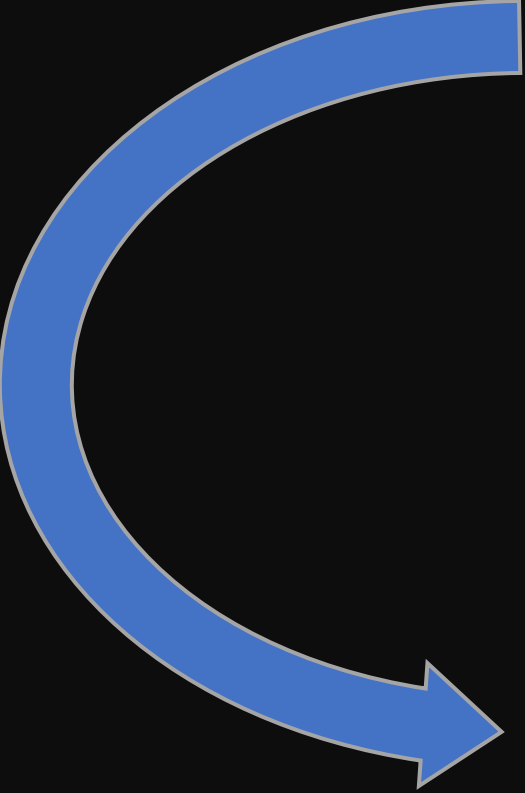
Filters:    Language: Any ⌄    Type: Any ⌄    Name: Any ⌄    Status: Any ⌄    ▽ Filters

| ⌄ Name ⌄ | Accuracy ⌄ | Status ⌄ | Type ⌄ | Language ⌄ | Created by ⌄ |
|---|---|---|---|---|---|
| ⌄ **Published (134)** | | | | | |
| ☐ Actuary reports ↗ | - | Ready to use | Built-In | English | Microsoft |
| ☐ Agreements ↗ | - | Ready to use | Built-In | English | Microsoft |
| ☐ Asset Management ↗ | - | Ready to use | Built-In | English | Microsoft |
| ☐ Bank statement ↗ | - | Ready to use | Built-In | English | Microsoft |
| ☐ Bank statement ↗ | - | Ready to use | Built-In | German | Microsoft |

# Exact Data Match Classifiers

| StudentID | First name | Last name | Date of Birth |
|-----------|------------|-----------|---------------|
| 0019-3947 | Ron | Weasley | 01-03-1980 |
| 0010-2394 | Harry | Potter | 31-06-1980 |
| 0015-5934 | Hermione | Granger | 19-09-1979 |

To: albus.dumbledore@hogwarts.com;

Cc:

Subject: Personal information of misbehaving student

Hello Albus,

I found one of our students misbehaving and brought him in for detention. Hereby I share with you his personal information.

StudentID: 0019-3947
Name: Ron Weasley
Date of Birth: 01-03-1980

Greetings,

Severus Snape

Microsoft Purview

New Microsoft Purview portal

Home

Compliance Manager

Data classification

Overview

Classifiers

Content explorer

Activity explorer

Data connectors

AI Hub (preview)

Alerts

Policies

Roles & scopes

Trials

Solutions

Catalog

App governance

# Home

What's new ?     + Add cards

Active alerts

## 0 active alert

| Alert name | Severity | Last activity |
|---|---|---|

# Preparation for next step

- Create the "EDM_DataUploaders" security group

- Install the Exact Data Match (EDM) agent

- Separate the hashing and uploading process

- Automate the process of uploading live data

Create exact data match sensitive information type in the New Experience workflow

Microsoft Purview

New Microsoft Purview portal

Home

Compliance Manager

Data classification

Overview

Classifiers

Content explorer

Activity explorer

Data connectors

AI Hub (preview)

Alerts

Policies

Roles & scopes

Trials

Solutions

Catalog

App governance

# Classifiers

Trainable classifiers    Sensitive info types    EDM classifiers

**New EDM experience**   On

ⓘ Learn the end-to-end workflow    Industry-specific sample files

[Why are there two experiences?](#)

Exact data match (EDM) classifiers use exact values from your org's data to detect matches instead of generic patterns. They can then be included in several compliance solutions to classify and protect sensitive data. [Learn more about EDM](#)

＋ Create EDM classifier

1 item        🔍 Search

| Name ↑ | Created by | Status |
|--------|-----------|--------|
| ☐ **StudentRecord** | Helgering | Source file not uploaded yet How to upload |

# Waiting….. (maximum of 1 hour)

Sign in

https://compliance.microsoft.com/dataclassificationclassifiers?viewid=exactdatamatch&tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

Microsoft Purview

New Microsoft Purview portal

Home

Compliance Manager

Data classification

Overview

Classifiers

Content explorer

Activity explorer

Data connectors

AI Hub (preview)

Alerts

Policies

Roles & scopes

Trials

Solutions

Catalog

# Classifiers

Trainable classifiers    Sensitive info types    **EDM classifiers**

**New EDM experience**    ◉ On

ⓘ Learn the end-to-end workflow    🗔 Industry-specific sample files

**Why are there two experiences?**

Exact data match (EDM) classifiers use exact values from your org's data to detect matches instead of generic patterns. They can then be included in several compliance solutions to classify and protect sensitive data. Learn more about EDM

＋ **Create EDM classifier**    1 item    🔍 Search

| Name ↑ | Created by | Status |
|---|---|---|
| ☐ **StudentRecord** | Helgering | Index complete |

# Goal achieved

- Identify how many and where sensitive information is stored with the Content Explorer

- Classify data that contain a StudentID with a Custom SIT

- Classify student information forms with a Fingerprint SIT

- Classify data that contain a potion recipe with a Trainable Classifier

- Classify data containing information that exactly matches the student record database with an Exact Data Match Classifier

# Takeaways

- Let classifiers be part of your classification process

- Don't rely on automation alone; involve your users

- Use a classifier that fits the type of data

- The task of managing classifiers is never done

in/myronhelgering     myronhelgering.com

# Q&A

Are there any questions?

# Thank you!

in/myronhelgering    myronhelgering.com