

Automated Data Security: Identifying Sensitive Data with Microsoft Purview



Level: Intermediate











Cybersecurity & Ransomware Live

Session Survey

- Your feedback is very important to us
- Please take a moment to complete the session survey found in the mobile app
- Use the QR code or search for "Converge360 Events" in your app store
- Find this session on the Agenda tab
- Click "Session Evaluation"
- Thank you!



About me

- Solution Lead @ Pink Elephant
- Microsoft Security MVP
- Blogger and Speaker
- Fantasy geek



Myron Helgering











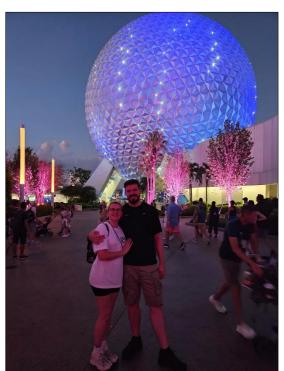




I 💘 Orlando

2019 2024 2025





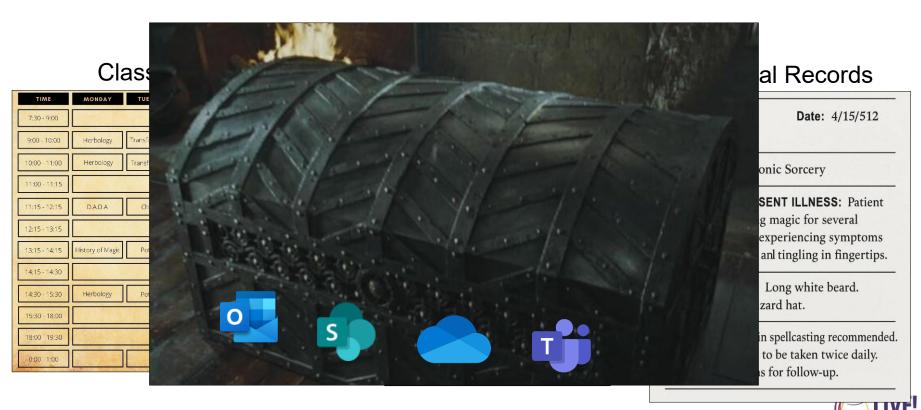


Agenda

01	Introduction	04	Trainable Classifiers
02	Sensitive Information Types	05	Exact Data Match Classifiers
03	Fingerprint SITs	06	Classifier Usage, Takeaways & Q&A

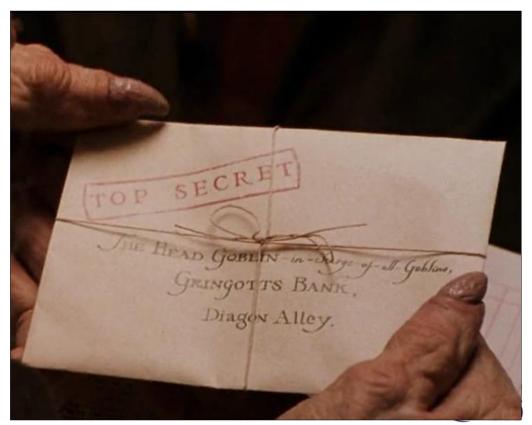


Why is classifying your sensitive data important?



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 sensitive data?

Manually









Meet Minerva McGonagall



Professor at Hogwarts School

Chief Information Security
 Officer (CISO)

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



License and permissions

- Access Data Explorer
- Apply protection automatically based on classifiers
- \longrightarrow
- M365 E5 or E5 add-on license

 Use advanced classifiers (e.g. trainable and EDM classifiers)

- Create classifiers and policies
- View content in Data Explorer

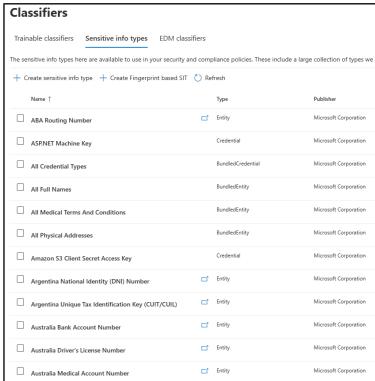
- Compliance Administrator role
- Content Explorer Content viewer



Built-in Sensitive Information Types

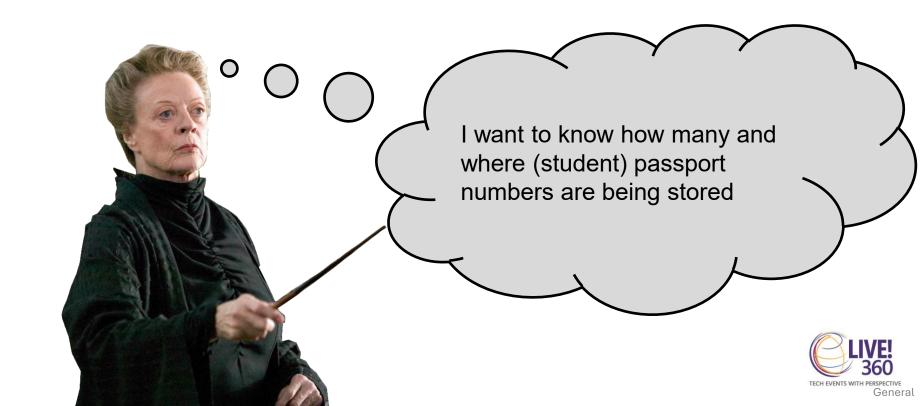
311 Built-in SITs to classify specific sensitive information;

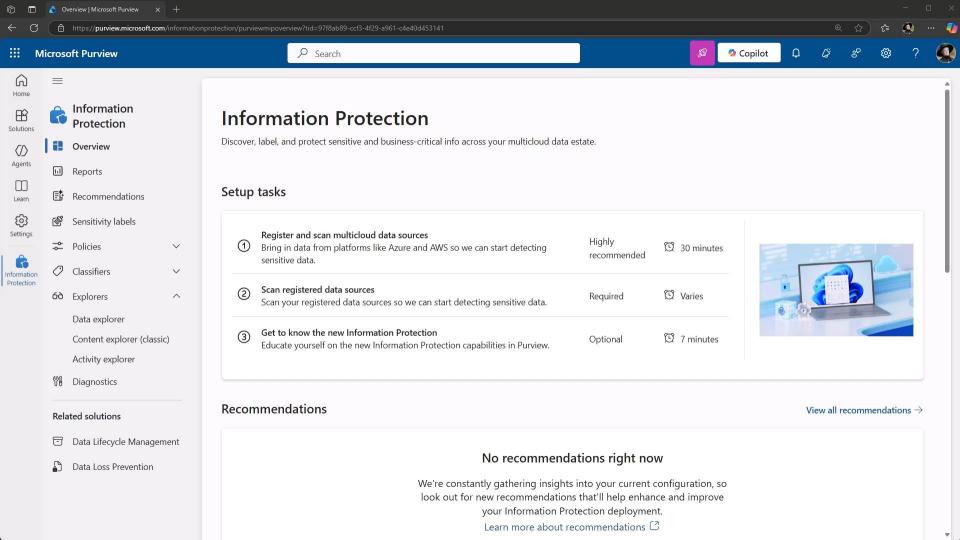
- Credit card
- Passport number
- National Insurance number (UK)
- SSN number (USA)
- Physical address
- Ip-address
- Driver's license number
- Password





Demo: Sensitive Information Types





Custom Sensitive Information Types

StudentID numbers

- Number format is 1234-5678
- Wordlist contains StudentID

Doc 1: Harry has **ID** number 389-9468

Doc 2: Harry has **StudentID** number 0397-3845

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



Lesson 1½: The 123s

Characters include normal letters, but digits as well. In fact, numbers 0-9 are also just characters and if you look at an ASCII table, they are listed sequentially.

Over the various lessons, you will be introduced to a number of special metacharacters used in regular expressions that can be used to match a specific type of character. In this case, the character \d can be used in place of any digit from 0 to 9. The preceding slash distinguishes it from the simple d character and indicates that it is a metacharacter.

Below are a few more lines of text containing digits. Try writing a pattern that matches all the digits in the strings below, and notice how your pattern matches anywhere within the string, not just starting at the first character. We will learn how to control this in a later lesson.

Exercise 1½: Matching Digits			
Task	Text		
Match	abc123xyz	②	
Match	define "123"	⊘	
Match	var g = 123;	⊘	
123		Continue >	

Lesson Notes

abc... Letters

123.... **Diaits**

١d Any Digit \D

Any Non-digit character

Any Character Period

Only a, b, or c

Not a, b, nor c [^abc]

[a-z] Characters a to z

Numbers 0 to 9

Any Alphanumeric character \w \W Any Non-alphanumeric character

m Repetitions m to n Repetitions

Zero or more repetitions

One or more repetitions

Optional character \s

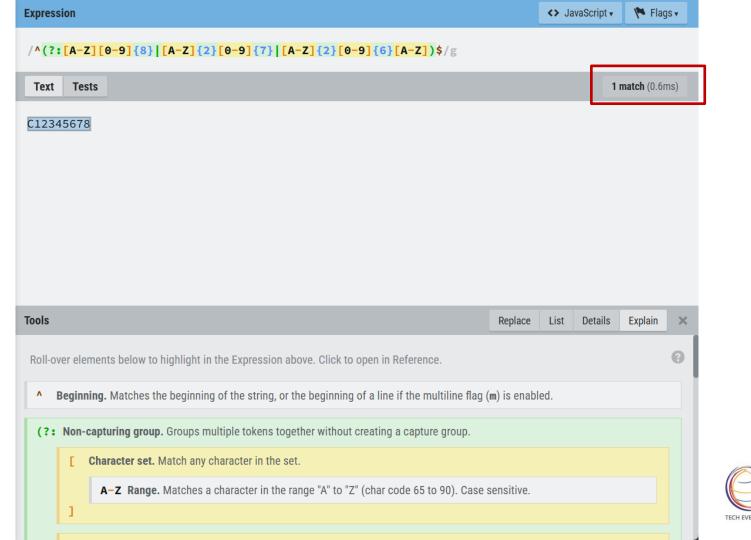
Any Whitespace Any Non-whitespace character

۸...\$ Capture Group

Capture Sub-group Capture all

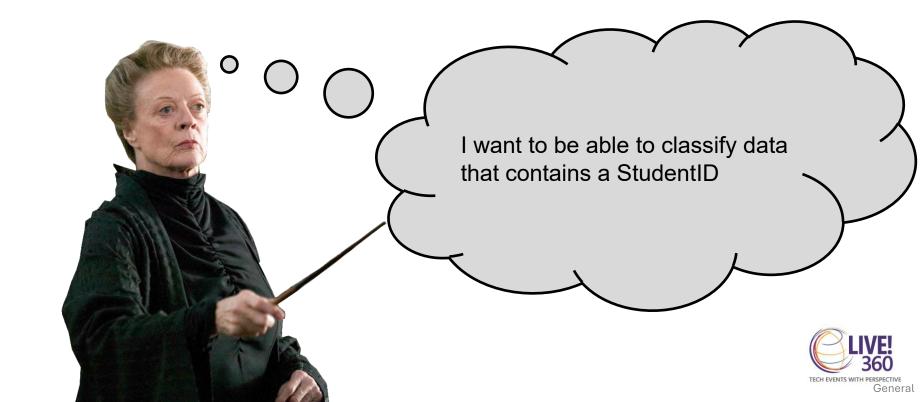
(abc|def) Matches abc or def

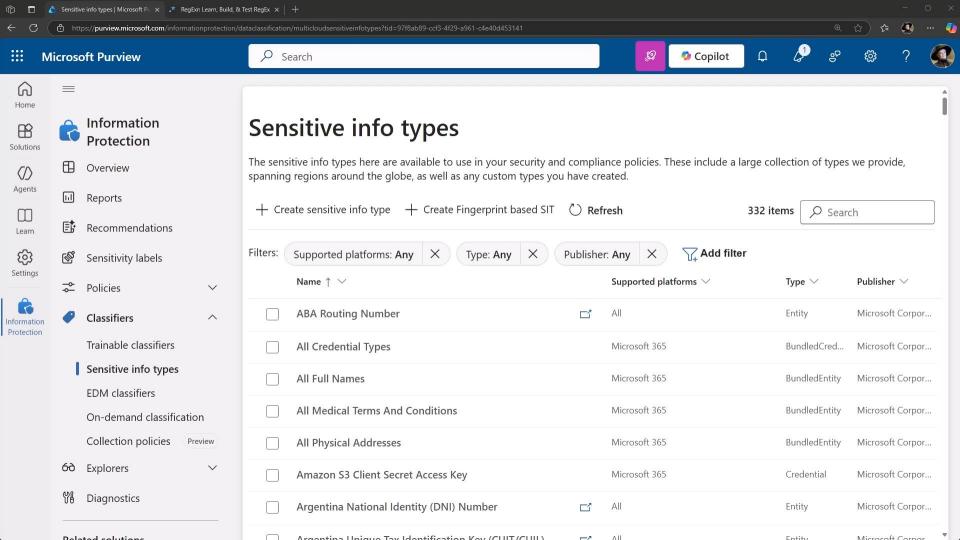
Solve the above task to continue on to the next problem, or read the Solution.

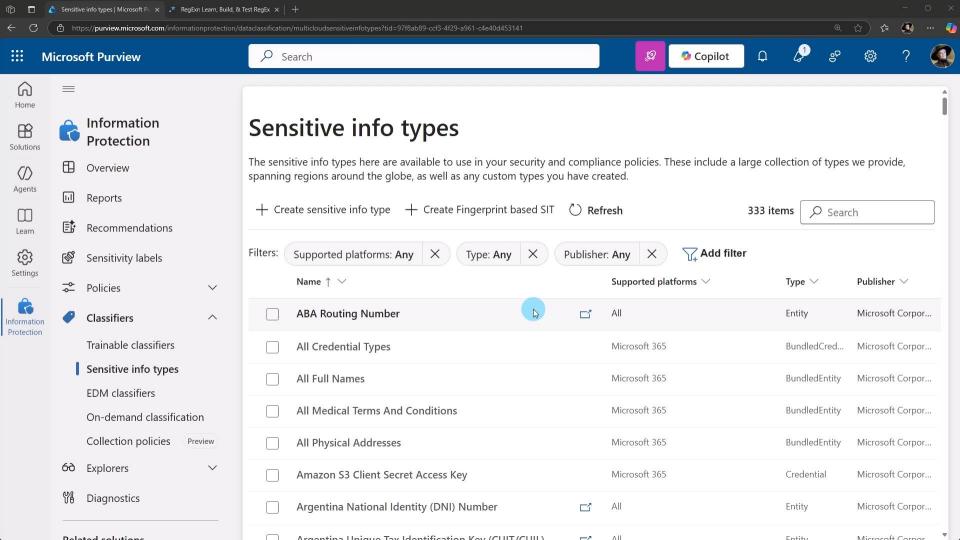


General

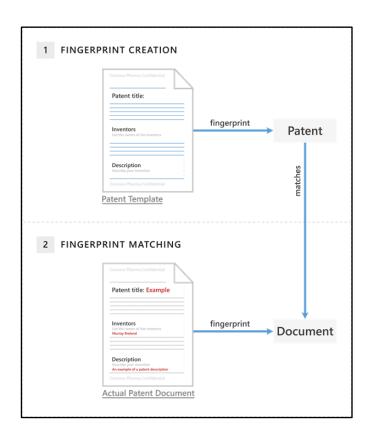
Demo: Custom Sensitive Information Types







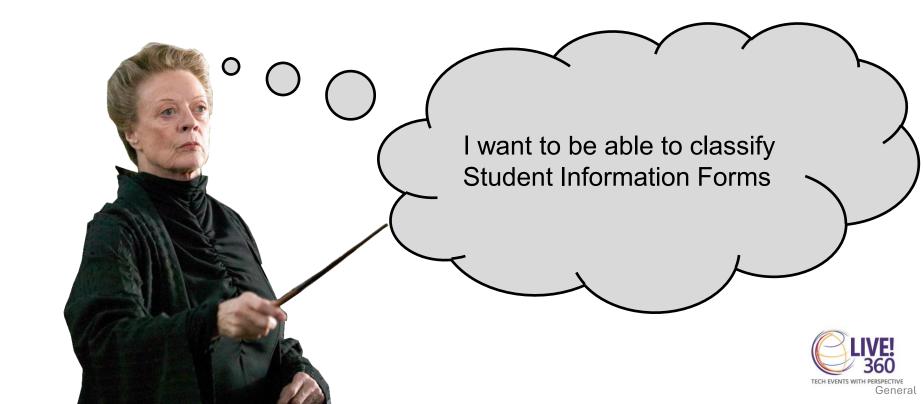
Fingerprint Sensitive Information Types

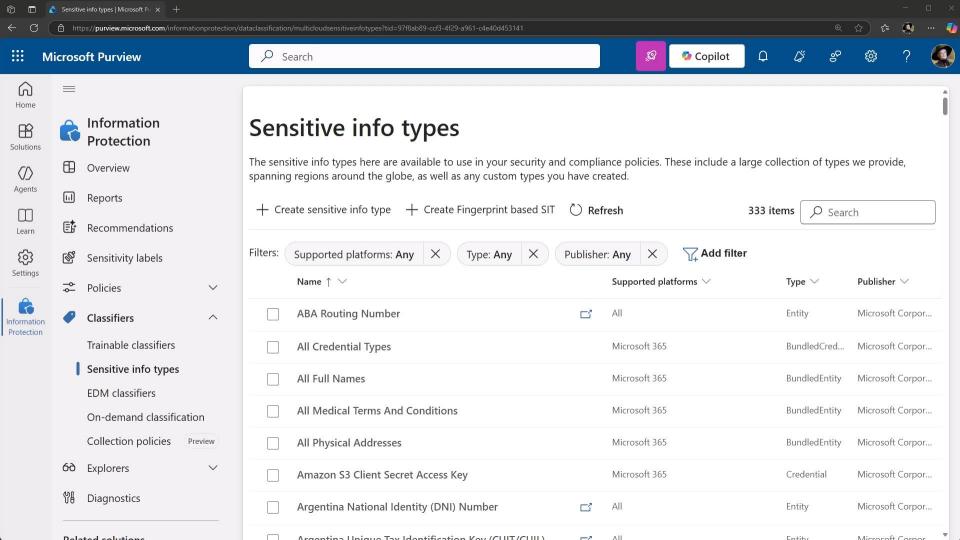


- Employee information forms
- Customer information forms
- Medical records
- Contracts
- Invoices
- Non-Disclosure Agreements (NDA)
- Or any other custom template file



Demo: Fingerprint Sensitive Information Types





Built-in Trainable Classifiers

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



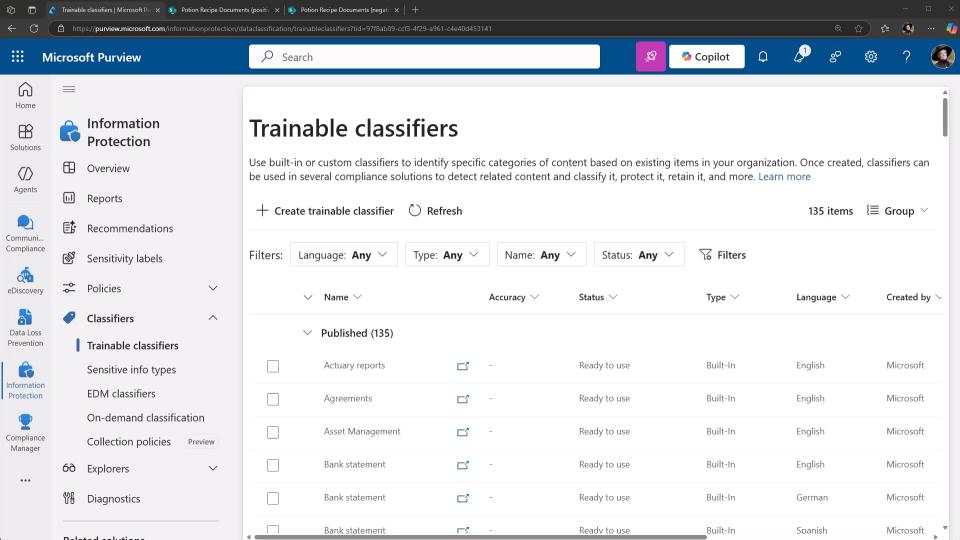
Custom Trainable Classifiers

- 1. 50-500 positive and negative samples (English only)
- 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

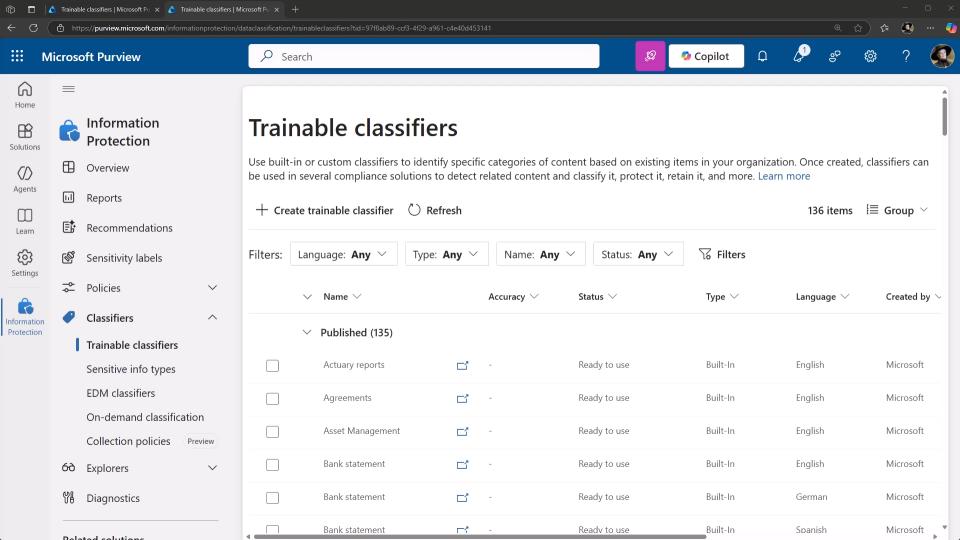




Waiting..... (maximum of 48 hours)







Exact Data Match Classifiers

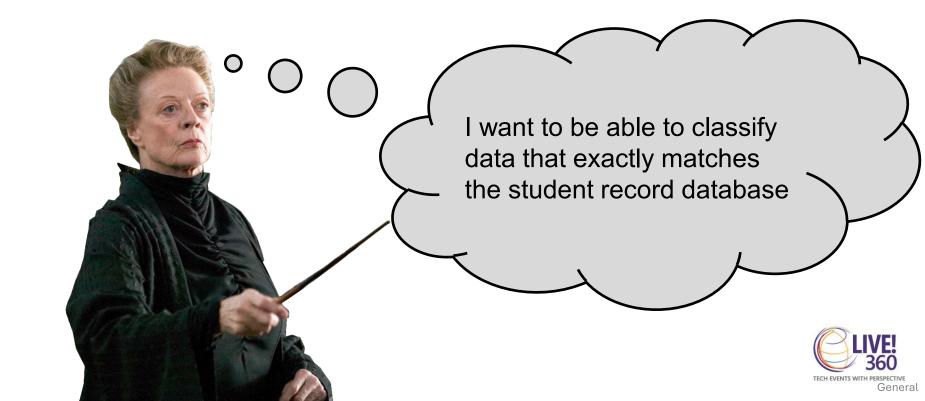


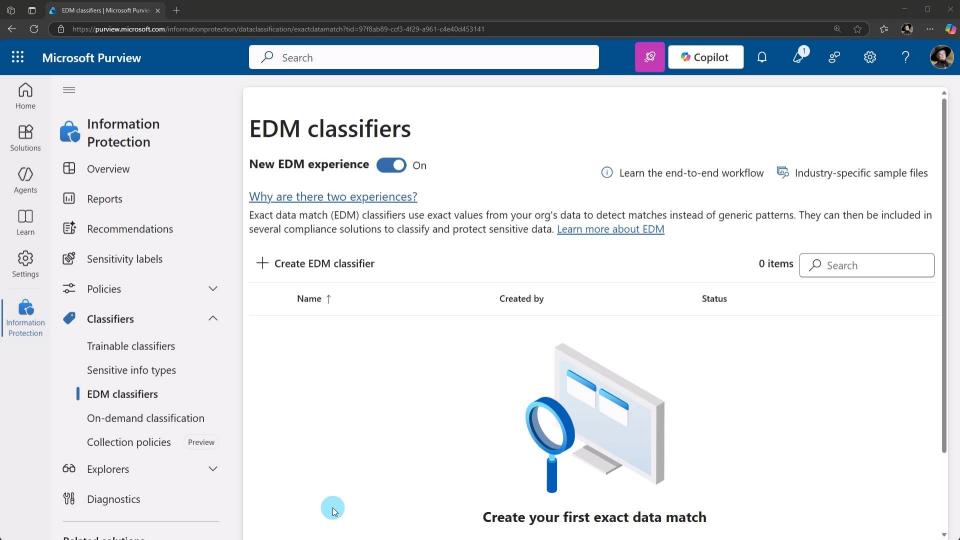
StudentID	First name	Last name	Date of Birth
0019-3947	Ron	Weasley	01-03-1980
0010-2394	Harry	Potter	31-06-1980
0015-5934	Hermione	Granger	19-09-1979

	То	albus.dumbledore@hogwarts.com;			
Send	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					
Greetir	ngs,				
Severus Snape					



Demo: Exact Data Match Classifiers



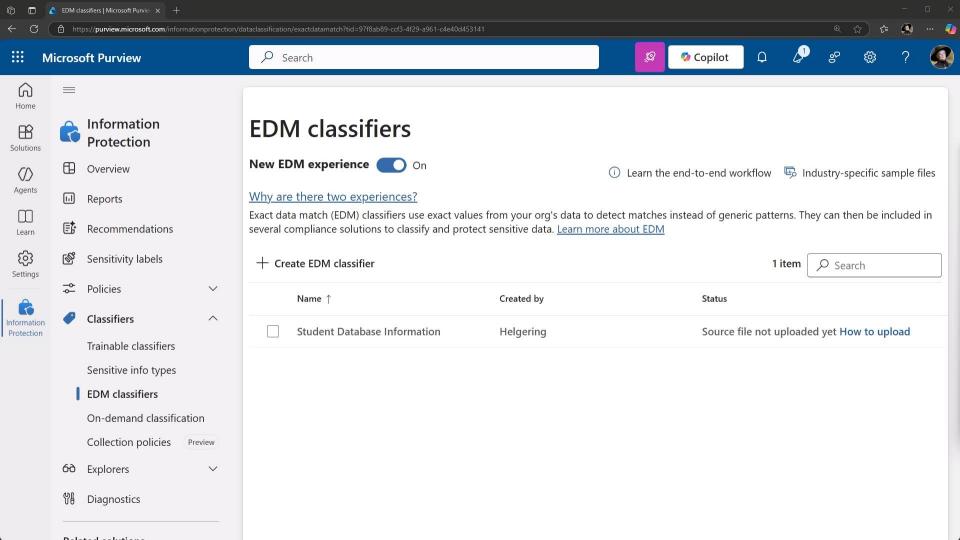


Preparation for next step

- 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

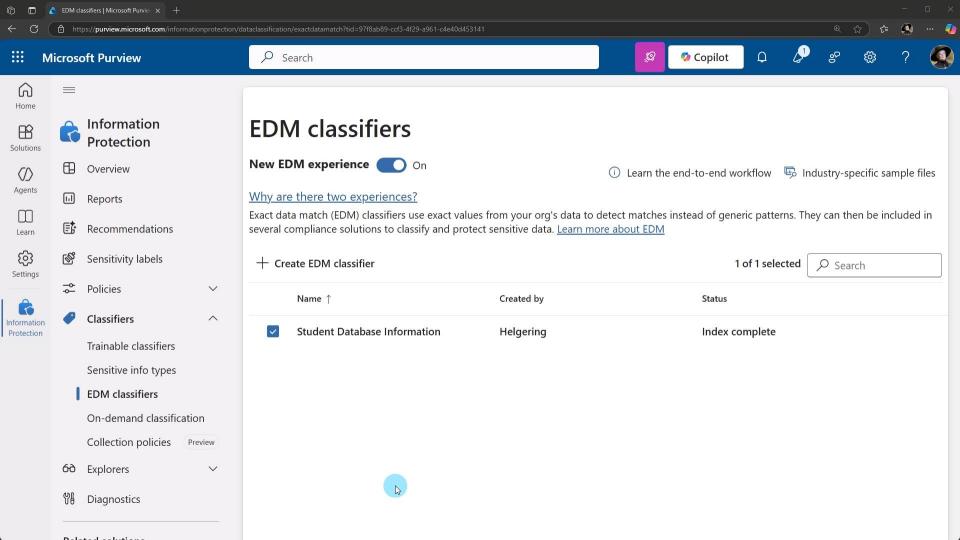




Waiting..... (maximum of 1 hour)







Goal achieved

- 1. Identify where sensitive information is stored with the Data Explorer
- 2. Classify data that contains a StudentID with a Custom SIT
- 3. Classify student information forms with a Fingerprint SIT
- 4. Classify potion recipes with a Trainable Classifier
- 5. Classify data containing information that exactly matches the student record database with an Exact Data Match Classifier



Where can we use Classifiers in Purview?

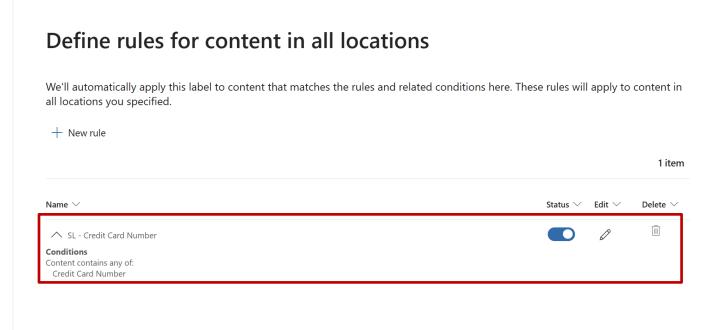




Information Protection

Auto-labeling policies > New policy

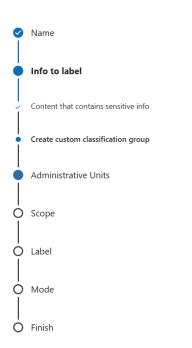






Data Lifecycle Management

Auto-labeling > Create auto-labeling policy



Define content that contains sensitive info

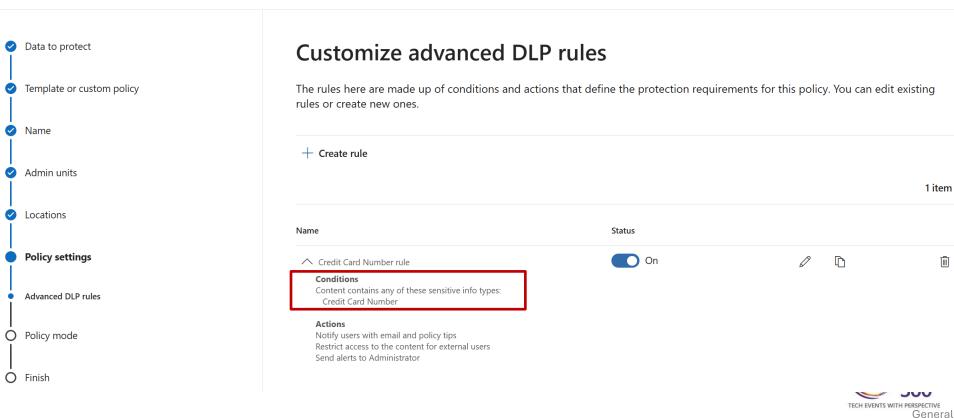
Choose a category of industry regulations to see the classification groups you can use to classify that information or create a custom group.

○ Content contains					
Group name * Credit Card Number				Group operator Any of these	Î
Sensitive info types					
Credit Card Number	High confidence	· i	Instance count 1	to Any i	Û
Add ✓					
ోన్లో Create group					
+ Add condition ∨					



Data Loss Prevention

Data loss prevention > Create policy



1 item

Create file policy Policy template * No template Policy name ¹ Policy severity * Category * DLP Description Files matching all of the following Edit and preview results Select a filter V + Add a filter Apply to: all files Select user groups: all file owners Inspection method **Data Classification Service** Any of the following occur: Match if Tredit Card Number ★ I Advanced settings

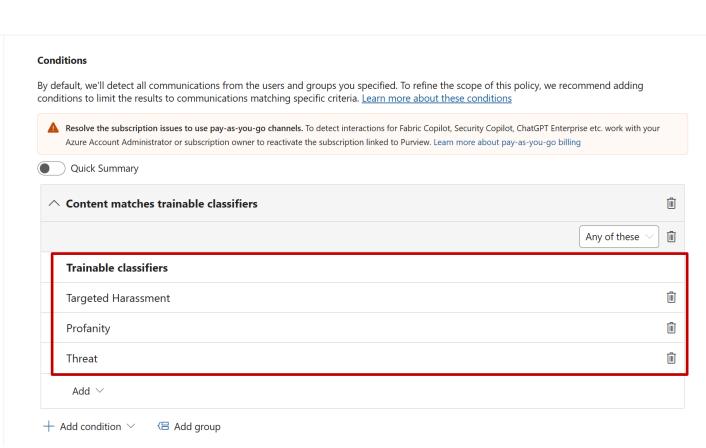
Defender for Cloud Apps



Communication Compliance

Communication compliance > New policy



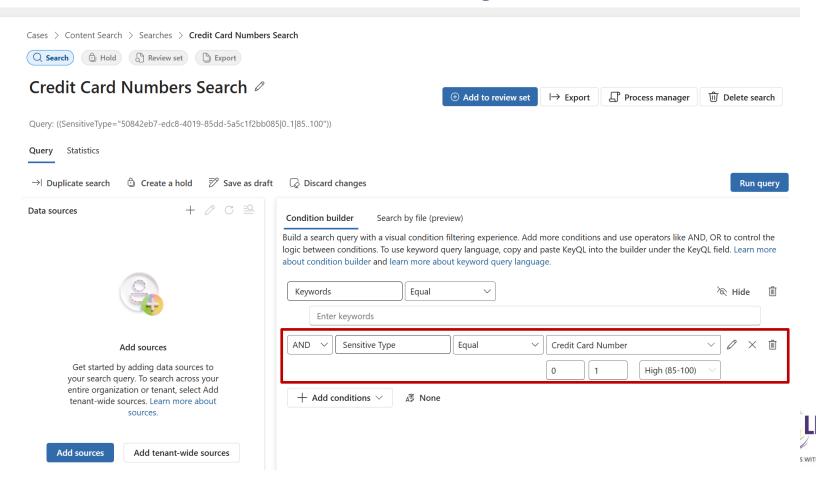


Insider Risk Management

Insider risk management > New insider risk policy Policy template Sensitive info types to prioritize Name and description Any activity associated with content that contains this sensitive info will be assigned a higher risk score. Admin units + Add or edit sensitive info types 1 item Users and groups Info type ∨ Credit Card Number X Content to prioritize Sensitive info types Scoring Triggering event O Indicators O Finish

General

eDiscovery



General

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







Time for questions!



Other sessions @ Live! 360

- Wednesday, November 19, 2025
 - Automated Data Security: Identifying Sensitive
 Data with Microsoft Purview (9:30am 10:45am)
- Thursday, November 20, 2025
 - Six Methods to Protect your Business from the Threat of Unmanaged Devices (1:00pm – 2:15pm)
 - The Challenge of Providing Secure Access to External Admins (2:30pm – 3:45pm)



Session Survey

- Your feedback is very important to us
- Please take a moment to complete the session survey found in the mobile app
- Use the QR code or search for "Converge360 Events" in your app store
- Find this session on the Agenda tab
- Click "Session Evaluation"
- Thank you!



Thank you!

