

Automating Data Security: Identifying Sensitive Data with Microsoft Purview

07:45 – 08:35



Myron Helgering



What are we going to discuss?



1. Introduction

2. Sensitive Information Types

3. Fingerprint SITs

4. Trainable Classifiers

5. Exact Data Match Classifiers

6. Takeaways and Q&A

About me

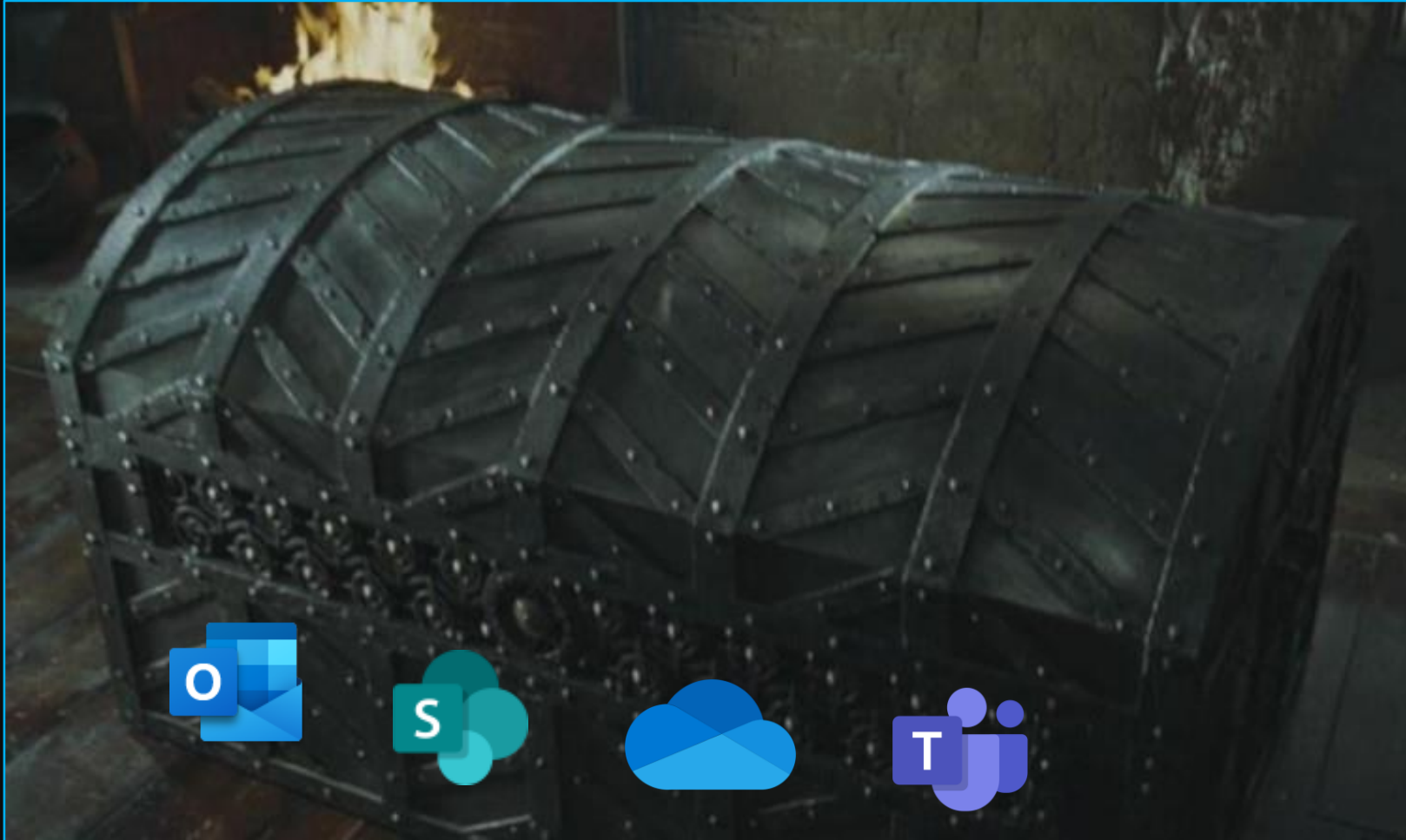
- CTO @ Cloud Life
- Microsoft Security MVP
- Blogger and Speaker
- Fantasy geek



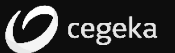
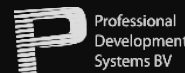
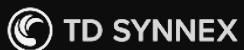
Myron Helgering

 [in/myronhelgering](https://www.linkedin.com/in/myronhelgering)  myronhelgering.com

Why is classifying your data important?



Microsoft



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



Microsoft



TD SYNEX



infor



DELL
Technologies



Professional
Development
Systems BV



INTERSTELLAR



kpn
Partner Network



INS PARK



cegeka

Who will classify data?

Manually (People)



Automatically (Magic)





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.



License and permissions

- Access Data Explorer
 - Apply protection automatically based on classifiers 
 - Create classifiers and policies 
 - View content in Data Explorer
- M365 E5 or E5 add-on license
 - Compliance Administrator role
 - Content Explorer Content viewer



Microsoft



TD SYNEX



Capgemini



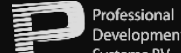
inforcer



DELL
Technologies



nerdio



Professional
Development
Systems BV



INTERSTELLAR



kpn
Partner Network



INS PARK



cegeka

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
- Password

Classifiers		
Trainable classifiers	<u>Sensitive info types</u>	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 p		
+ Create sensitive info type + Create Fingerprint based SIT Refresh		
Name ↑	Type	Publisher
<input type="checkbox"/> ABA Routing Number	Entity	Microsoft Corporation
<input type="checkbox"/> ASP.NET Machine Key	Credential	Microsoft Corporation
<input type="checkbox"/> All Credential Types	BundledCredential	Microsoft Corporation
<input type="checkbox"/> All Full Names	BundledEntity	Microsoft Corporation
<input type="checkbox"/> All Medical Terms And Conditions	BundledEntity	Microsoft Corporation
<input type="checkbox"/> All Physical Addresses	BundledEntity	Microsoft Corporation
<input type="checkbox"/> Amazon S3 Client Secret Access Key	Credential	Microsoft Corporation
<input type="checkbox"/> Argentina National Identity (DNI) Number	Entity	Microsoft Corporation
<input type="checkbox"/> Argentina Unique Tax Identification Key (CUIT/CUIL)	Entity	Microsoft Corporation
<input type="checkbox"/> Australia Bank Account Number	Entity	Microsoft Corporation
<input type="checkbox"/> Australia Driver's License Number	Entity	Microsoft Corporation
<input type="checkbox"/> Australia Medical Account Number	Entity	Microsoft Corporation



Microsoft



TD SYNEX



Capgemini



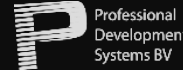
inforcer



DELL
Technologies



nerdio



Professional
Development
Systems BV



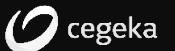
INTERSTELLAR



kpn
Partner Network



INSPARK



cegeka



Demo: Sensitive Information Types



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

Overview | Microsoft Purview

https://purview.microsoft.com/informationprotection/purviewmipoverview?tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

Microsoft Purview

Search

Copilot

Home

Solutions

Agents

Learn

Settings

Information Protection

Information Protection

Overview

Reports

Recommendations

Sensitivity labels

Policies

Classifiers

Explorers

Data explorer

Content explorer (classic)

Activity explorer

Diagnostics

Related solutions

Data Lifecycle Management

Data Loss Prevention

Information Protection

Discover, label, and protect sensitive and business-critical info across your multicloud data estate.

Setup tasks

1	Register and scan multicloud data sources Bring in data from platforms like Azure and AWS so we can start detecting sensitive data.	Highly recommended	30 minutes
2	Scan registered data sources Scan your registered data sources so we can start detecting sensitive data.	Required	Varies
3	Get to know the new Information Protection Educate yourself on the new Information Protection capabilities in Purview.	Optional	7 minutes

View all recommendations →

No recommendations right now

We're constantly gathering insights into your current configuration, so look out for new recommendations that'll help enhance and improve your Information Protection deployment.

Learn more about recommendations



Custom Sensitive Information Types

StudentID numbers

- Number format is 1234-5678
- Wordlist includes 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/>



Microsoft



TD SYNEX



Capgemini

infor



DELL
Technologies



nerdio



Professional
Development
Systems BV



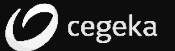
INTERSTELLAR



kpn
Partner Network



INS PARK



cegeka

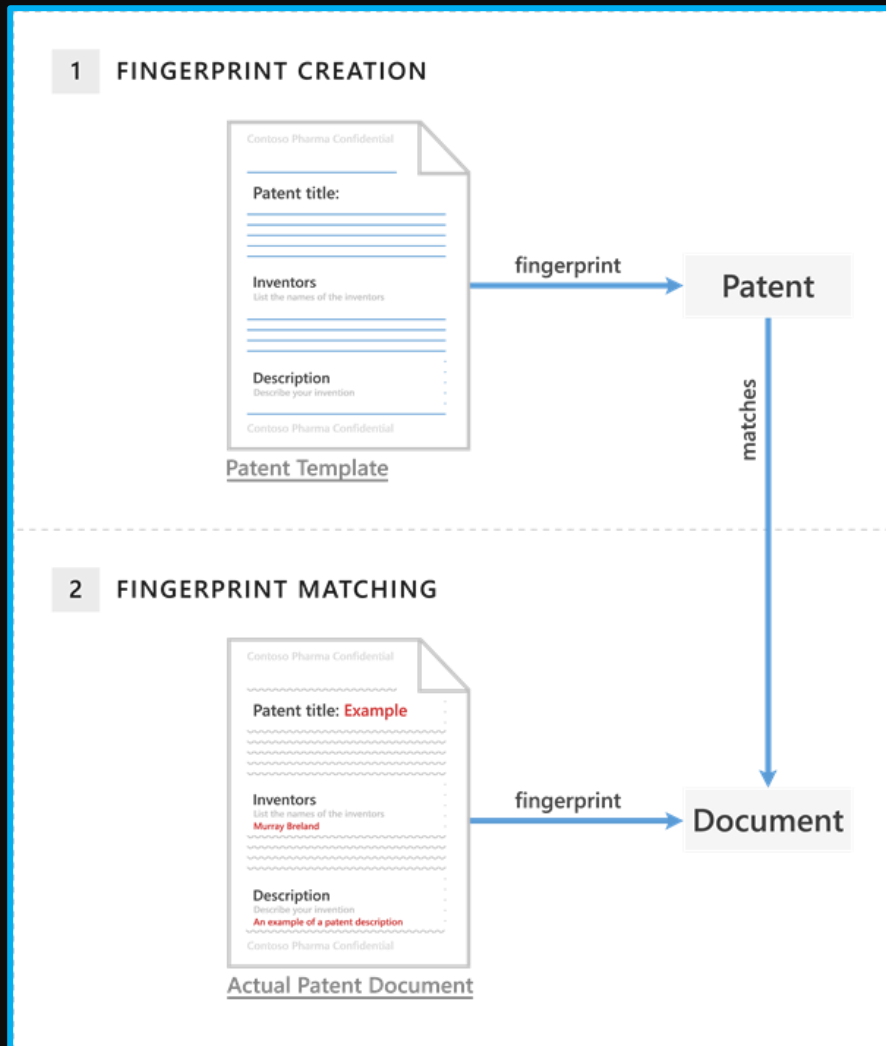


Demo: Custom Sensitive Information Types



**I want to be able to classify
data that contains a StudentID**

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



TD SYNEX



Capgemini



inforcer



DELL
Technologies



nerdio



Professional
Development
Systems BV



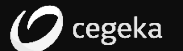
INTERSTELLAR



kpn
Partner Network



INS PARK



cegeka



Demo: Fingerprint Sensitive Information Types



**I want to be able to classify
Student Information Forms**

Home

Solutions

Agents

Learn

Settings

Information Protection

Information Protection

Overview

Reports

Recommendations

Sensitivity labels

Policies

Classifiers

Trainable classifiers

Sensitive info types

EDM classifiers

On-demand classification

Collection policies

Preview

Explorers

Diagnostics

Related solutions

Microsoft Purview

Search

Copilot

Sensitive info types

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

333 items

Search

Filters:

Supported platforms: Any

Type: Any

Publisher: Any

Add filter

	Name	Supported platforms	Type	Publisher
<input type="checkbox"/>	ABA Routing Number	All	Entity	Microsoft Corpor...
<input type="checkbox"/>	All Credential Types	Microsoft 365	BundledCred...	Microsoft Corpor...
<input type="checkbox"/>	All Full Names	Microsoft 365	BundledEntity	Microsoft Corpor...
<input type="checkbox"/>	All Medical Terms And Conditions	Microsoft 365	BundledEntity	Microsoft Corpor...
<input type="checkbox"/>	All Physical Addresses	Microsoft 365	BundledEntity	Microsoft Corpor...
<input type="checkbox"/>	Amazon S3 Client Secret Access Key	Microsoft 365	Credential	Microsoft Corpor...
<input type="checkbox"/>	Argentina National Identity (DNI) Number	All	Entity	Microsoft Corpor...
<input type="checkbox"/>	Argentina Unique Tax Identification Key (CUIT/CUIL)	All	Entity	Microsoft Corpor...



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 (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



**I want to be able to
classify potion recipes**

Trainable classifiers | Microsoft Pu x

Potion Recipe Documents (positiv x

Potion Recipe Documents (negati x

+

https://purview.microsoft.com/informationprotection/dataclassification/trainableclassifiers?tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

Microsoft Purview

Search

Copilot

1

?

Home

Solutions

Agents

Communi... Compliance

eDiscovery

Data Loss Prevention

Information Protection

Compliance Manager

...

Information Protection

Overview

Reports

Recommendations

Sensitivity labels

Policies

Classifiers

Trainable classifiers

Sensitive info types

EDM classifiers

On-demand classification

Collection policies

Preview

Explorers

Diagnostics

Related solutions

Trainable 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 (135)						
<input type="checkbox"/>	Actuary reports		-	Ready to use	Built-In	English	Microsoft
<input type="checkbox"/>	Agreements		-	Ready to use	Built-In	English	Microsoft
<input type="checkbox"/>	Asset Management		-	Ready to use	Built-In	English	Microsoft
<input type="checkbox"/>	Bank statement		-	Ready to use	Built-In	English	Microsoft
<input type="checkbox"/>	Bank statement		-	Ready to use	Built-In	German	Microsoft
<input type="checkbox"/>	Bank statement		-	Ready to use	Built-In	Spanish	Microsoft

Waiting.... (maximum of 48 hours)



Microsoft



TD SYNEX

Capgemini

infor

DELL
Technologies

nerdio

P
Professional
Development
Systems BV

INTERSTELLAR



kpn
Partner Network

INS PARK



cegeka

Trainable classifiers | Microsoft Pu x

Trainable classifiers | Microsoft Pu x

+

←

↺

https://purview.microsoft.com/informationprotection/dataclassification/trainableclassifiers?tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

🔍

☆

☆

👤

...

🌐

Microsoft Purview

🔍

Search

🚀

Copilot

🔔

🔔1

🗣️

⚙️

?

👤

🏠

Home

🧩

Solutions

👤

Agents

📖

Learn

⚙️

Settings

🔒

Information Protection

☰

Information Protection

📊

Overview

📈

Reports

📌

Recommendations

🔍

Sensitivity labels

🔗

Policies

▼

🏷️

Classifiers

▲

📑

Trainable classifiers

🔍

Sensitive info types

🔍

EDM classifiers

🔍

On-demand classification

🔍

Collection policies

Preview

🔍

Explorers

▼

🔍

Diagnostics

Trainable 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

136 items

📑

Group

▼

Filters:

Language: Any

Type: Any

Name: Any


Status: Any

🔍


Filters

	Name	Accuracy	Status	Type	Language	Created by	
▼	Published (135)						
<input type="checkbox"/>	Actuary reports	🔗	-	Ready to use	Built-In	English	Microsoft
<input type="checkbox"/>	Agreements	🔗	-	Ready to use	Built-In	English	Microsoft
<input type="checkbox"/>	Asset Management	🔗	-	Ready to use	Built-In	English	Microsoft
<input type="checkbox"/>	Bank statement	🔗	-	Ready to use	Built-In	English	Microsoft
<input type="checkbox"/>	Bank statement	🔗	-	Ready to use	Built-In	German	Microsoft
<input type="checkbox"/>	Bank statement	🔗	-	Ready to use	Built-In	Spanish	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



TD SYNEX



Capgemini



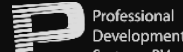
inforcer



DELL
Technologies



nerdio



Professional
Development
Systems BV



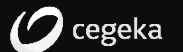
INTERSTELLAR



kpn
Partner Network



INS PARK



cegeka



Demo: Exact Data Match Classifiers



I want to be able to classify data that exactly matches the student record database

EDM classifiers | Microsoft Purview

https://purview.microsoft.com/informationprotection/dataclassification/exactdatamatch?tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

Microsoft Purview

Search

Copilot

1

?

Home

Solutions

Agents

Learn

Settings

Information Protection

Information Protection

Overview

Reports

Recommendations

Sensitivity labels

Policies

Classifiers

Trainable classifiers

Sensitive info types

EDM classifiers

On-demand classification

Collection policies

Preview

Explorers

Diagnostics

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

0 items

Search

Name ↑	Created by	Status
--------	------------	--------

Create your first exact data match



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



TD SYNnex



Capgemini



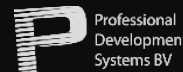
infor



Dell
Technologies



nerdio



Professional
Development
Systems BV



INTERSTELLAR



kpn
Partner Network



INSPIRE



cegeka

EDM classifiers | Microsoft Purview

https://purview.microsoft.com/informationprotection/dataclassification/exactdatamatch?tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

Microsoft Purview

Search

Copilot

1

?

Home

Solutions

Agents

Learn

Settings

Information Protection

Information Protection

Overview

Reports

Recommendations

Sensitivity labels

Policies

Classifiers

Trainable classifiers

Sensitive info types

EDM classifiers

On-demand classification

Collection policies

Explorers

Diagnostics

Related solutions

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
<div><input type="checkbox"/></div> Student Database Information	Helgering	Source file not uploaded yet How to upload

Waiting.... (maximum of 1 hour)



Microsoft



TD SYNEX

Capgemini

infor

DELL
Technologies

nerdio



Professional
Development
Systems BV

INTERSTELLAR



kpn
Partner Network



INS PARK



cegeka

EDM classifiers | Microsoft Purview

https://purview.microsoft.com/informationprotection/dataclassification/exactdatamatch?tid=97f8ab89-ccf3-4f29-a961-c4e40d453141

Microsoft Purview

Search

Copilot

1

?

Home

Solutions

Agents

Learn

Settings

Information Protection

Information Protection

Overview

Reports

Recommendations

Sensitivity labels

Policies

Classifiers

Trainable classifiers

Sensitive info types

EDM classifiers

On-demand classification

Collection policies

Explorers

Diagnostics

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 of 1 selected

Search

Name ↑	Created by	Status
<div><div></div>Student Database Information</div>	Helgering	Index complete



Goal achieved

1. Identify where sensitive information is stored with the **Data Explorer**
2. Classify data that contain 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**



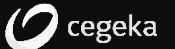
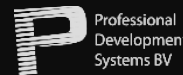
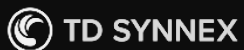


Takeaways

- 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](https://www.linkedin.com/in/myronhelgering)

 myronhelgering.com





Please evaluate this session in the App.

THANK YOU

Are there any questions?

