M365
SECURITY,
COMPLIANCE &
MANAGEMENT

Advanced Classification and Auto Labeling

Using Microsoft Information Protection to detect and protect sensitive data

Microsoft Information Protection

Tony Themelis (Product Management)

Nir Hendler (Customer Adoption Team)

Adam Bell (Customer Adoption Team)

Microsoft Information Protection (MIP)

BUILT-IN

Built-in labeling and protection experience in Microsoft 365 apps, Microsoft 365 services, other MS services like Power BI, Edge and Windows

INTELLIGENT

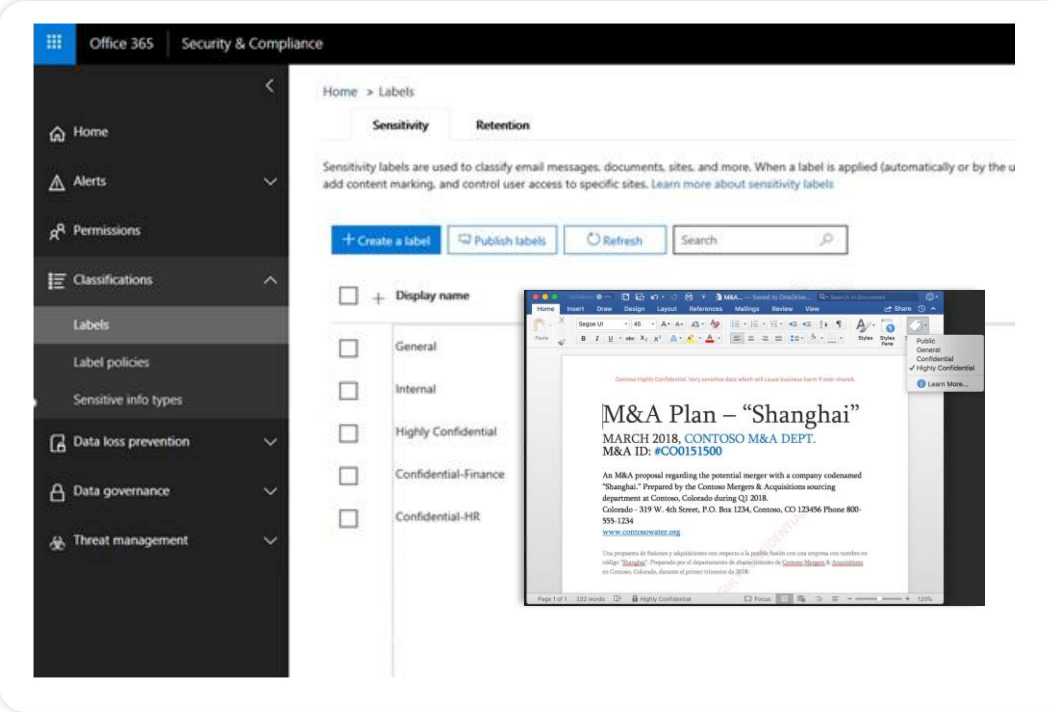
Accuracy in classification via ML based trainable classifiers, exact data match and entities

UNIFIED

Single admin console to configure and manage your policies and view analytics across on-premises, Microsoft 365 apps, Microsoft 365 services, 3rd party services and Windows devices

EXTENSIBLE

MIP platform extends the protection experience, in a consistent way, to popular non-Microsoft apps and services



Sensitivity labels are used to classify email messages, documents, sites, and more. When a label is applied (automatically or by the u add content marking, and control user access to specific sites. Learn more about sensitivity labels

	Display name
□	General
□	Internal
□	Highly Confidential
□	Confidential-Finance
□	Confidential-HR

Contoso Highly Confidential. Very sensitive data which will cause business harm if over-shared.

M&A Plan – "Shanghai"

MARCH 2018, CONTOSO M&A DEPT.
M&A ID: #CO0151500

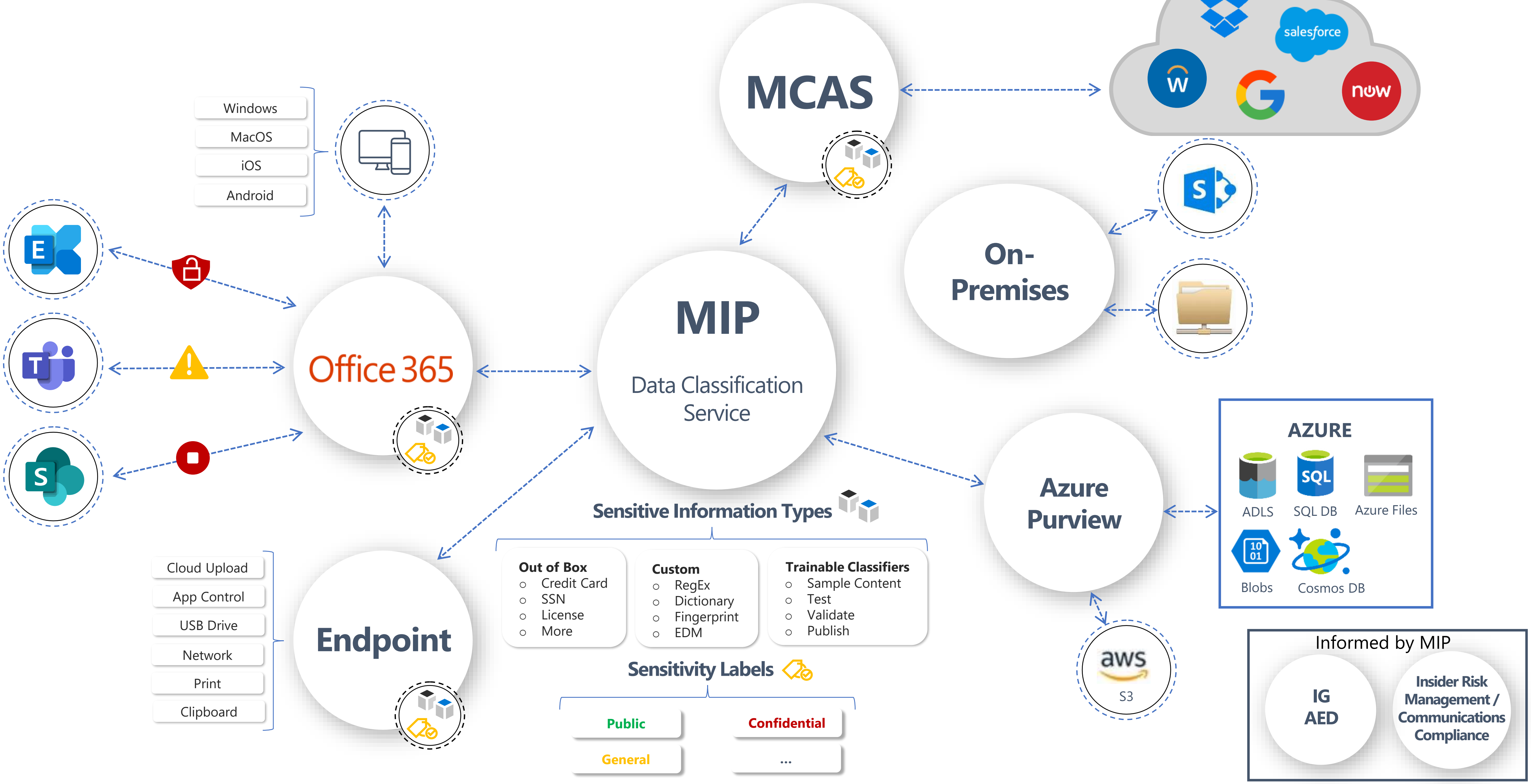
An M&A proposal regarding the potential merger with a company codenamed "Shanghai." Prepared by the Contoso Mergers & Acquisitions sourcing department at Contoso, Colorado during Q1 2018.
Colorado - 319 W. 4th Street, P.O. Box 1234, Contoso, CO 123456 Phone 800-555-1234
www.contosowater.org

Una propuesta de fusiones y adquisiciones con respecto a la posible fusión con una empresa con nombre en código "Shanghai". Preparado por el departamento de abastecimiento de Contoso Mergers & Acquisitions en Contoso, Colorado, durante el primer trimestre de 2018.

Microsoft Information Protection is a built-in, intelligent, unified and extensible platform and solution to protect sensitive data.

M365 SECURITY, COMPLIANCE & MANAGEMENT

Microsoft Information Protection



MCAS

Windows
MacOS
iOS
Android

Office 365

MIP

Data Classification Service

On-Premises

Azure Purview

AZURE
ADLS
SQL DB
Azure Files
Blobs
Cosmos DB

S3

Cloud Upload
App Control
USB Drive
Network
Print
Clipboard

Endpoint

Sensitive Information Types

Out of Box	Custom	Trainable Classifiers
Credit Card	RegEx	Sample Content
SSN	Dictionary	Test
License	Fingerprint	Validate
More	EDM	Publish

Sensitivity Labels

Public	Confidential
General	...

Informed by MIP
IG AED
Insider Risk Management / Communications Compliance

MIP policy scope and coverage

Best practice: Think across all environments

Office apps across platforms:

Label and protect Office files natively across Windows, Mac, iOS, Android and Web Clients

SharePoint Online:

Automatically label and protect sensitive files in SharePoint Online and OneDrive for Business

SharePoint sites teams, Office 365 groups:

Label and protect sensitive SharePoint Sites, Teams, Office 365 Groups, Power BI artifacts

Exchange Online:

Automatically label and protect sensitive emails in Exchange Online

On-prem:

Classify and label data in on-prem repositories

Non-Microsoft clouds and SaaS apps:

Extend protection through Microsoft Cloud App Security to third party clouds and SaaS apps

Unified Label Management in Microsoft 365 Compliance center

Auto labeling for sensitive files and emails

Auto labeling is a native Microsoft service that runs in SharePoint Online, OneDrive for Business and Exchange Online

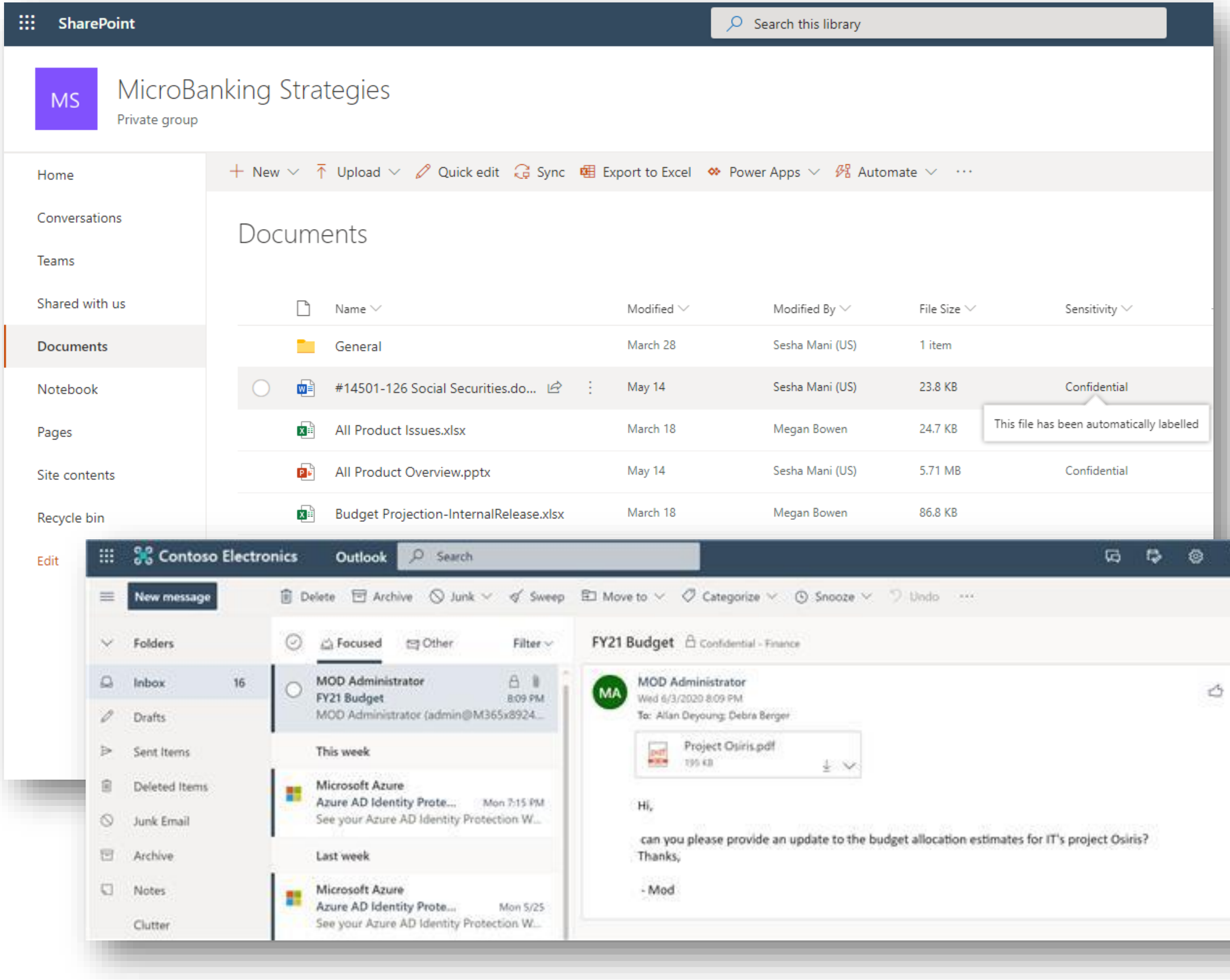
Sensitive files are automatically detected and labeled at rest

Sensitive emails are automatically detected and labeled in transit

Configure policies using regulatory templates, 200+ out of box sensitive info or custom types, named entities, Exact Data Match and ML models

New: increased scale supports scoping policy to all locations and faster simulation results (coming Q3)

AKA.MS/Autoclassification



Documents

Name	Modified	Modified By	File Size	Sensitivity
General	March 28	Sesha Mani (US)	1 item	
#14501-126 Social Securities.do...	May 14	Sesha Mani (US)	23.8 KB	Confidential
All Product Issues.xlsx	March 18	Megan Bowen	24.7 KB	This file has been automatically labelled
All Product Overview.pptx	May 14	Sesha Mani (US)	5.71 MB	Confidential
Budget Projection-InternalRelease.xlsx	March 18	Megan Bowen	86.8 KB	

MOD Administrator
FY21 Budget 8:09 PM
MOD Administrator (admin@M365x8924...

This week

Microsoft Azure
Azure AD Identity Prote... Mon 7:15 PM
See your Azure AD Identity Protection W...

Last week

Microsoft Azure
Azure AD Identity Prote... Mon 5/25
See your Azure AD Identity Protection W...

FY21 Budget Confidential - Finance

MA MOD Administrator
Wed 6/3/2020 8:09 PM
To: Allan Deyoung; Debra Berger

Project Osiris.pdf
195 KB

Hi,

can you please provide an update to the budget allocation estimates for IT's project Osiris?
Thanks,

- Mod

Understand what auto labeling is intended for

Client-side auto labeling		Service-side auto labeling
Auto	Recommended	Auto
• Triggers off sensitive content found in files, emails • Part of the label definition • Works in interactive (data-in-use) scenarios • Policy tips inform the user of policy verdicts • Covers Office clients, on-prem scanner		• Triggers off sensitive content found in files, emails • Defined in an auto labeling policy • Works for data-at-rest and data-in-motion • Covers OneDrive, SharePoint and Exchange • Includes simulation mode • Prefer service-side auto labeling over MCAS for OneDrive and SharePoint

Consider default labels vs. auto labeling

- Default labels:
  - Intended to apply the same label to any unlabeled files and emails, independent of content
  - Applied interactively when users create or edit documents and emails
- Auto labeling:
  - Intended to detect sensitive content in files and emails, and apply the relevant label
  - Applied at rest to files and in motion to emails

Auto labeling feature flow

1

Pick your scope

- Option 1: ALL – SharePoint sites, OneDrive accounts and Email users
- Option 2: Subset of sites or accounts – can use PowerShell for longer lists
- *Roadmap: OneDrive groups will be supported by end of year*

2

Simulate in your production environment

- Simulation is fast – it normally takes a few hours to run depending on the size of your tenant
- Simulation is not intrusive – no labels are applied
- Insights are best achieved on real production data

3

Gain confidence in your auto labeling policy

- Iterate and experiment

4

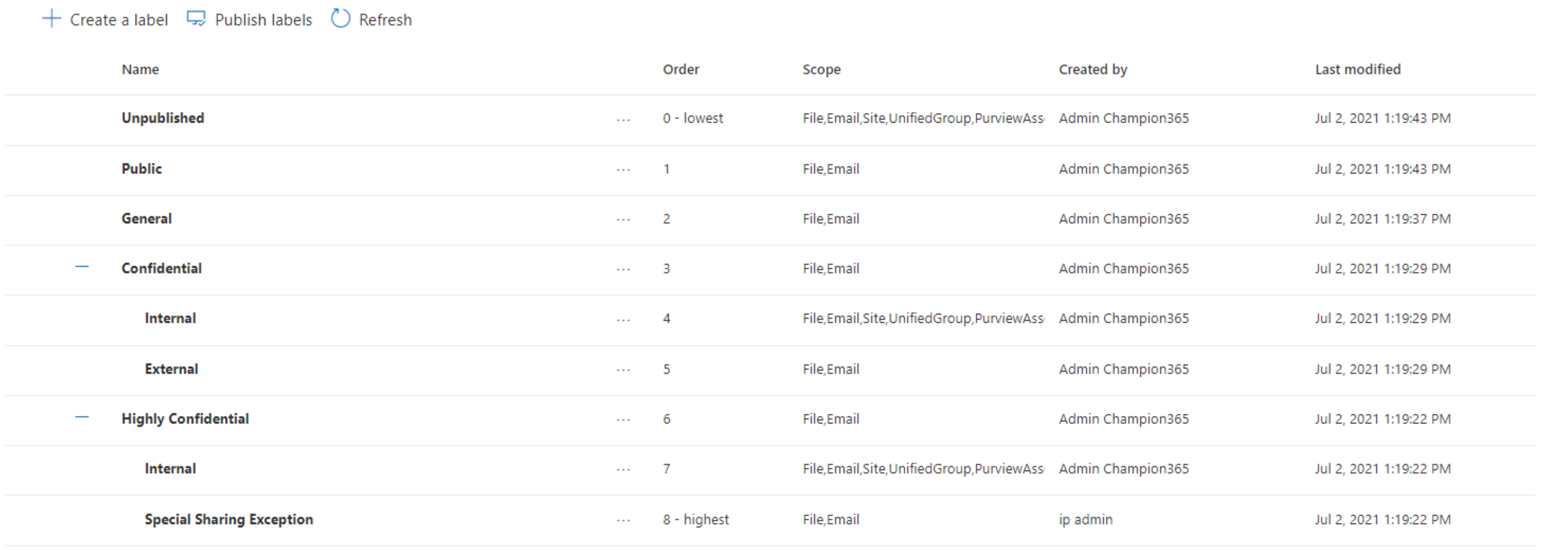
Enforce auto labeling policies after validating simulation results

- Existing Office Files at rest (Word, Excel, PowerPoint) in OneDrive & SharePoint are automatically labeled
- New files added after the policy is enforced are also labeled
- Emails in transit are automatically scanned for sensitive information and labeled
- *New: Auto labeling for emails can also be triggered off contextual predicates*

M365 SECURITY, COMPLIANCE & MANAGEMENT

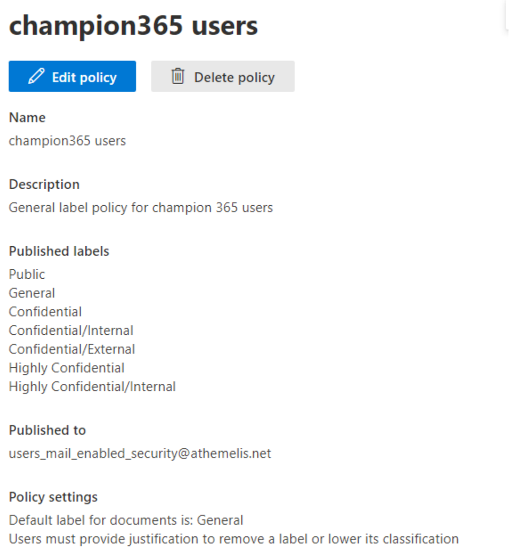
Before you start auto labeling – set up labels and label policy

1. Create sensitivity labels
   - Label taxonomy and hierarchy is defined
   - Label contain protection actions such as encryption



Name	Order	Scope	Created by	Last modified
Unpublished	0 - lowest	File,Email,Site,UnifiedGroup,PurviewAss	Admin Champion365	Jul 2, 2021 1:19:43 PM
Public	1	File,Email	Admin Champion365	Jul 2, 2021 1:19:43 PM
General	2	File,Email	Admin Champion365	Jul 2, 2021 1:19:37 PM
Confidential	3	File,Email	Admin Champion365	Jul 2, 2021 1:19:29 PM
Internal	4	File,Email,Site,UnifiedGroup,PurviewAss	Admin Champion365	Jul 2, 2021 1:19:29 PM
External	5	File,Email	Admin Champion365	Jul 2, 2021 1:19:29 PM
Highly Confidential	6	File,Email	Admin Champion365	Jul 2, 2021 1:19:22 PM
Internal	7	File,Email,Site,UnifiedGroup,PurviewAss	Admin Champion365	Jul 2, 2021 1:19:22 PM
Special Sharing Exception	8 - highest	File,Email	ip admin	Jul 2, 2021 1:19:22 PM

2. Publish label policy to users
   - Define default and mandatory labels
   - Remember default labels are applied here, not in auto labeling



champion365 users

Name
champion365 users

Description
General label policy for champion 365 users

Published labels
Public
General
Confidential
Confidential/Internal
Confidential/External
Highly Confidential
Highly Confidential/Internal

Published to
users_mail_enabled_security@athemelis.net

Policy settings
Default label for documents is: General
Users must provide justification to remove a label or lower its classification

Decide which classifiers to use in your auto labeling policy

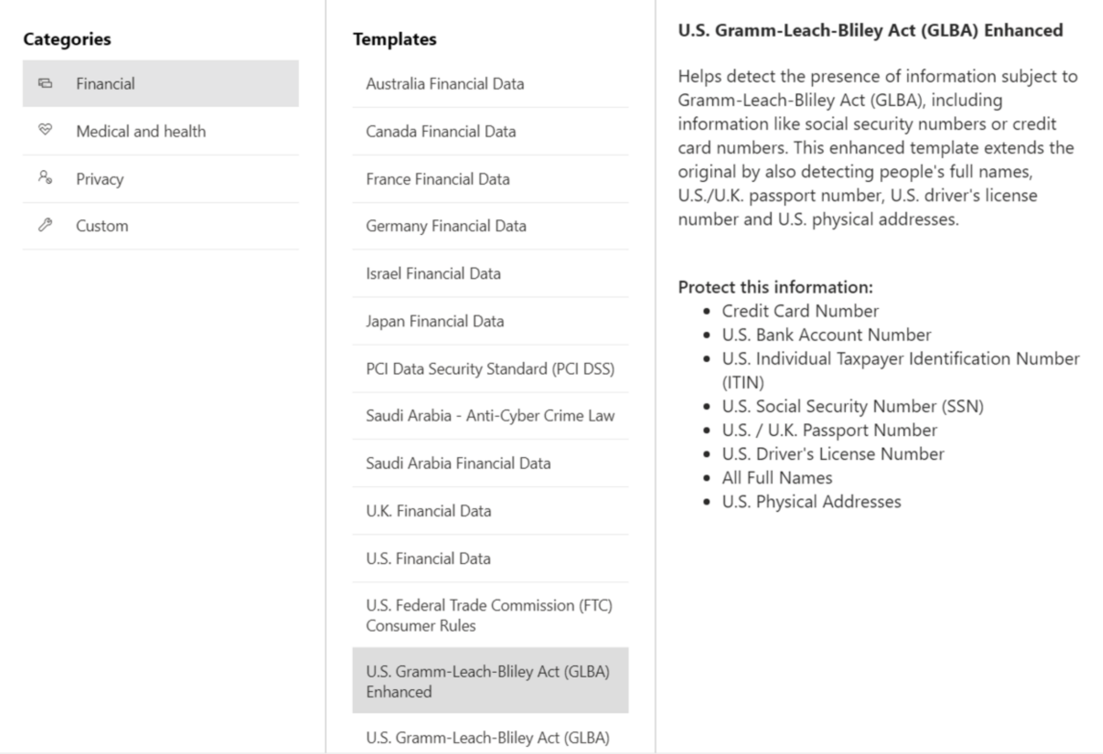
Sensitive Info Types	Named Entities	Exact Data Match	Trainable Classifiers
• 200+ out of the box info types like SSN, CCN • Can be cloned and edited • Create your own • Supports regex, keywords and dictionaries	• 50+ entities covering person name, medical terms and drug names • Best used in combination with sensitive info types	• Provides a lookup to exactly match content with unique customer data • Supports 100m rows and multiple lookup fields	• 10+ out of the box machine learning classifiers like resume, source code • Create your own classifier based on business data
Available today	*In private preview, available Q3*	*Available today*	*Available Q4*

Available templates provide pre-defined policies that use the above classifiers:

- Covering multiple industry and geographical regulatory requirements
- Easily customizable
- Can be edited to meet customer needs

Easily get started by simulating with a template

- Fine tune as necessary by running the simulation multiple times



Categories

- Financial
- Medical and health
- Privacy
- Custom

Templates

- Australia Financial Data
- Canada Financial Data
- France Financial Data
- Germany Financial Data
- Israel Financial Data
- Japan Financial Data
- PCI Data Security Standard (PCI DSS)
- Saudi Arabia - Anti-Cyber Crime Law
- Saudi Arabia Financial Data
- U.K. Financial Data
- U.S. Financial Data
- U.S. Federal Trade Commission (FTC) Consumer Rules
- U.S. Gramm-Leach-Bliley Act (GLBA) Enhanced
- U.S. Gramm-Leach-Bliley Act (GLBA)

U.S. Gramm-Leach-Bliley Act (GLBA) Enhanced

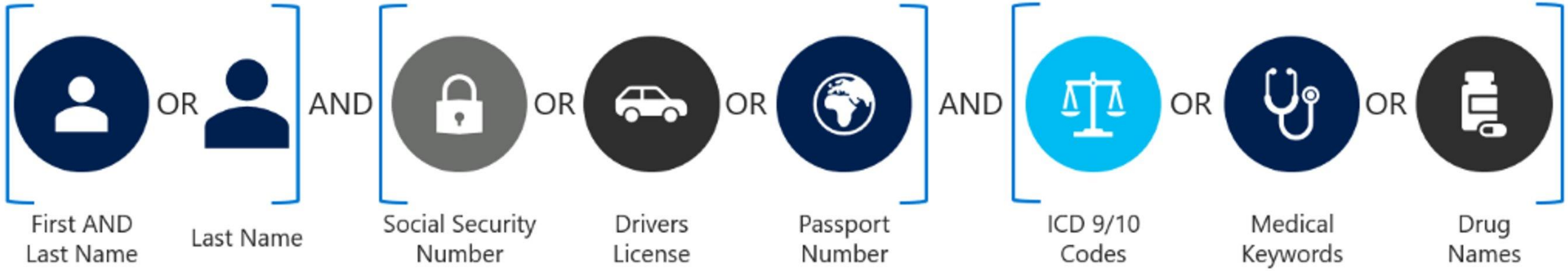
Helps detect the presence of information subject to Gramm-Leach-Bliley Act (GLBA), including information like social security numbers or credit card numbers. This enhanced template extends the original by also detecting people's full names, U.S./U.K. passport number, U.S. driver's license number and U.S. physical addresses.

Protect this information:

- Credit Card Number
- U.S. Bank Account Number
- U.S. Individual Taxpayer Identification Number (ITIN)
- U.S. Social Security Number (SSN)
- U.S. / U.K. Passport Number
- U.S. Driver's License Number
- All Full Names
- U.S. Physical Addresses

Iterate and experiment: picking and grouping of classifiers

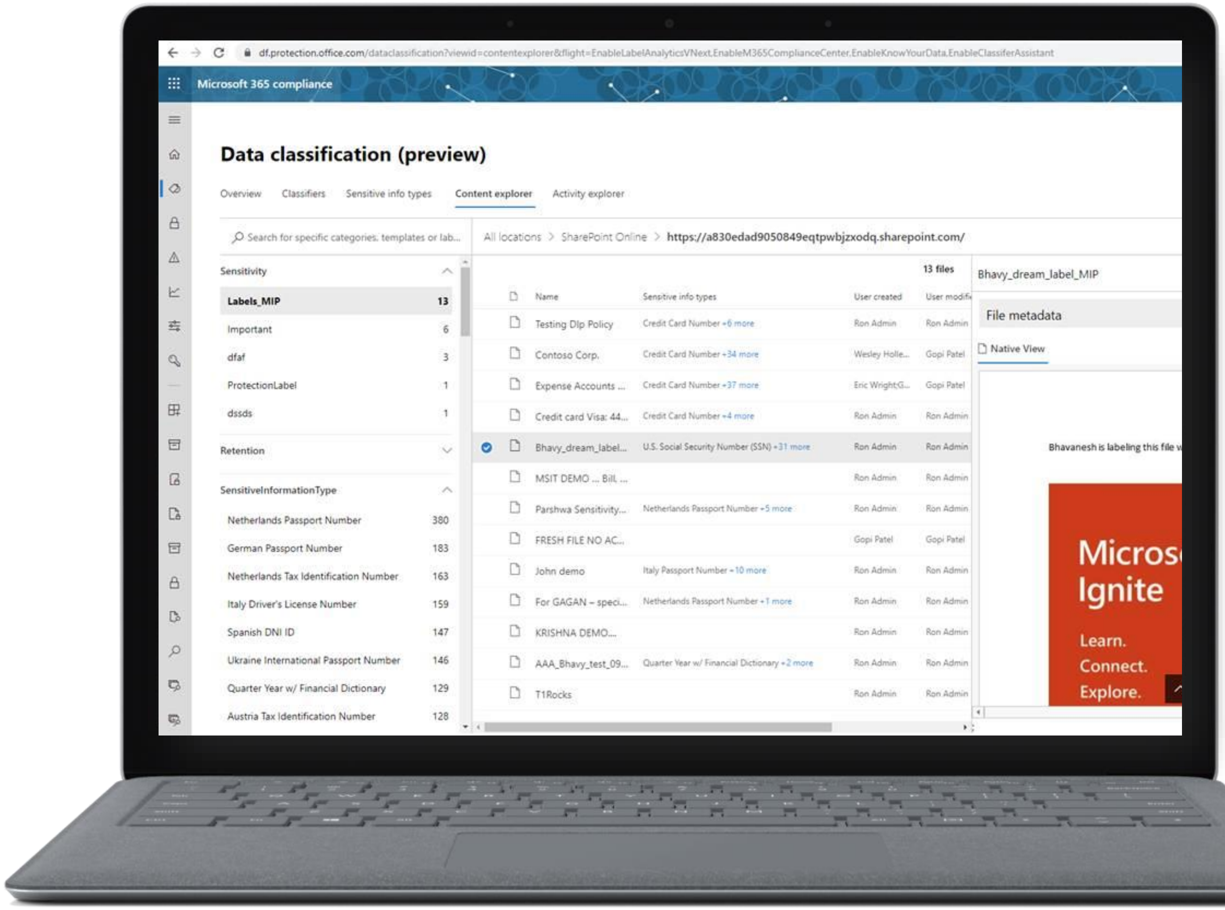
- Group like information types together
- Use Boolean operators to combine groups
- Understand confidence levels and how they are defined
  - Low confidence may be good!
- Use thresholds to determine severity
  - It’s okay to use different thresholds for individual classifiers



Use Content Explorer to gain insight into your data

A deeper view into your underlying data

- Visibility into amount of sensitive data in a document that triggered the classification to be applied
- Ability to filter by label to get a more detailed view including locations of where documents are stored
- Integrated native viewer displays a rich view of documents, providing context for policy creation



df.protection.office.com/dataclassification?viewid=contentexplorer&flight=EnableLabelAnalyticsVNext,EnableM365ComplianceCenter,EnableKnowYourData,EnableClassiferAssistant

Microsoft 365 compliance

Data classification (preview)

All locations > SharePoint Online > https://a830edad9050849eqtpwbjzxodq.sharepoint.com/

Sensitivity	
Labels_MIP	13
Important	6
dfaf	3
ProtectionLabel	1
dssds	1
Retention	
SensitiveInformationType	
Netherlands Passport Number	380
German Passport Number	183
Netherlands Tax Identification Number	163
Italy Driver's License Number	159
Spanish DNI ID	147
Ukraine International Passport Number	146
Quarter Year w/ Financial Dictionary	129
Austria Tax Identification Number	128

13 files

Name	Sensitive info types	User created	User modifi
Testing Dlp Policy	Credit Card Number +6 more	Ron Admin	Ron Admin
Contoso Corp.	Credit Card Number +34 more	Wesley Holle...	Gopi Patel
Expense Accounts ...	Credit Card Number +37 more	Eric Wright;G...	Gopi Patel
Credit card Visa: 44...	Credit Card Number +4 more	Ron Admin	Ron Admin
Bhavy_dream_label...	U.S. Social Security Number (SSN) +31 more	Ron Admin	Ron Admin
MSIT DEMO ... Bill, ...		Ron Admin	Ron Admin
Parshwa Sensitivity...	Netherlands Passport Number +5 more	Ron Admin	Ron Admin
FRESH FILE NO AC...		Gopi Patel	Gopi Patel
John demo	Italy Passport Number +10 more	Ron Admin	Ron Admin
For GAGAN – speci...	Netherlands Passport Number +1 more	Ron Admin	Ron Admin
KRISHNA DEMO...		Ron Admin	Ron Admin
AAA_Bhavy_test_09...	Quarter Year w/ Financial Dictionary +2 more	Ron Admin	Ron Admin
T1Rocks		Ron Admin	Ron Admin

Bhavy_dream_label_MIP

Bhavanesh is labeling this file w

Auto labeling considerations

- Simulation is fast
  - Scanning for sensitive content across all locations scoped in the auto labeling policy completes within a few hours
- Labeling in OneDrive and SharePoint:
  - Happens after simulation is done and the policy is set to enforce
  - Labels matched files at a rate of 25k per day
  - Data-in-motion takes precedence over data-at-rest
  - Limit of 1m matched files to be labeled per policy
  - *Roadmap: improved stats on auto label enforcement coming by end of year*
- Labeling in Exchange
  - Emails are labeled in transit as they are sent
- File type coverage
  - Modern office docs (docx, pptx, xlsx) are supported today
  - *Roadmap: PDF support targeted for next year (summer 2022)*

How to get started and involved

Auto labeling

- https://aka.ms/SPOAutoClassification

For other previews see:

- https://aka.ms/ODSPSecurityPreviews
- https://aka.ms/mip-preview

Additional resources

- https://aka.ms/SPOLabels
- https://aka.ms/SPOAutoClassification
- OSS – https://aka.ms/mipc/oss
- SBD YouTube –https://aka.ms/mipc/SBD

Thank you