

SIT - Quick Start (Sensitive Info Types)

Please consider this a "Work in Progress"



Sensitive Info Types



Where do SIT's fit in Data Governance

Information Protection

Information Governance Records Management Collaboration Governance

SIT's, Sensitivity Labels & Retention Labels

Exchange

SharePoint

Teams

Azure/M365 Security Platform & DLP Policies

Document title paate presenter of client harne

Data protection & challenges



"As an administrator, how can I ensure the organizational data is protected, and prevent access to unauthorized data?"

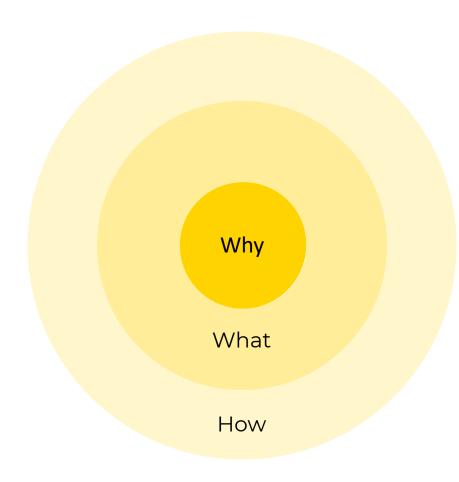
"As a data owner, how can I ensure the organizational data I work with is protected"?

"How can my productivity stay high while using those & ensure compliance?"

"As an **administrator**, how do I gain **oversight** of the sensitive data stored in my tenant, where, and by whom it's accessed?" "As an Auditor, how can I ensure that our Organization ensures we meet our privacy & regulatory requirements and are compliant?"

Is the business workflow/process's documented or still in tribal knowledge?

Before you start:



Why

- What is the business impact of what you are trying to protect?
- + Can you quantify the impact?
- + If not, is it worth the effort?

What

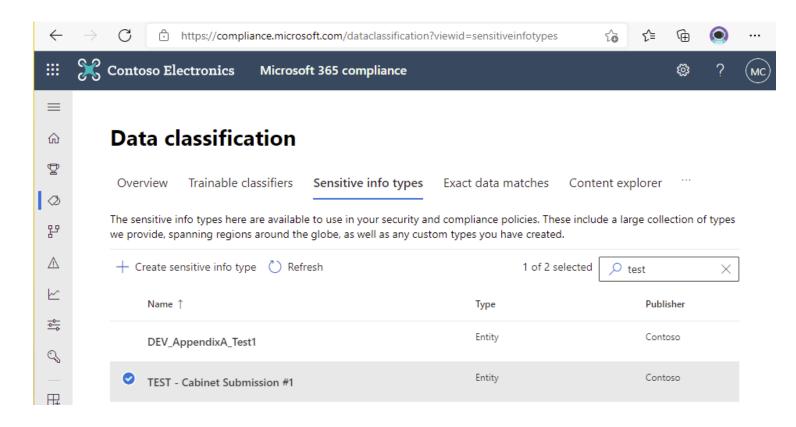
- + Do you have samples of what you are trying to protect
- Do you know how many times it currently occurs?
- + What is the impact on down the line processes? (If you create 100 events will this be 100 incidents?)

How

- Is there a current SIT that you can use as an example of what you are trying?
- Can you draw a flowchart of the logic you are applying?
- Now we can look at the technology to implement SITs

Data Classification can be broken down as:

- + Trainable Classifiers
- Sensitive Info Types (SIT's)
- Exact Data Matches (typically from DB)



Apart from Sensitive Info Types (SIT's) **Data Classification** can be broken down as:

- + Trainable Classifiers
- Sensitive Info Types (SIT's)
 - + 204+ Out of Box new and improved & can be copied/edited
- + Exact Data Matches (typically from DB)
- Doc Fingerprinting (exchange only)

See details in Speaker notes for summary + links

In this session we will be focusing on Sensitive Info Types (SIT's) only

Sensitive info types > Edit sensitive info type

Sensitive Info Types (SIT's):

- Copy an Out of Box SIT or create new
- Creating a SIT requires = Name & Description (naturally)

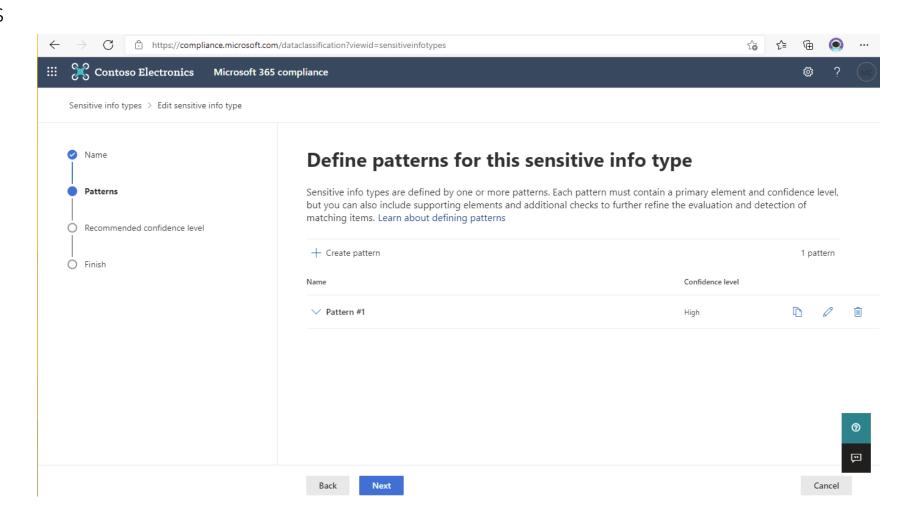
Name Name your sensitive info type This name and description will appear in compliance policies that support sensitive info Patterns types, so be sure to enter text that helps admins easily understand what info will be detected. Recommended confidence le... Name * Finish TEST - Cabinet Submission #1 Description * TEST - Cabinet Submission #1

Next

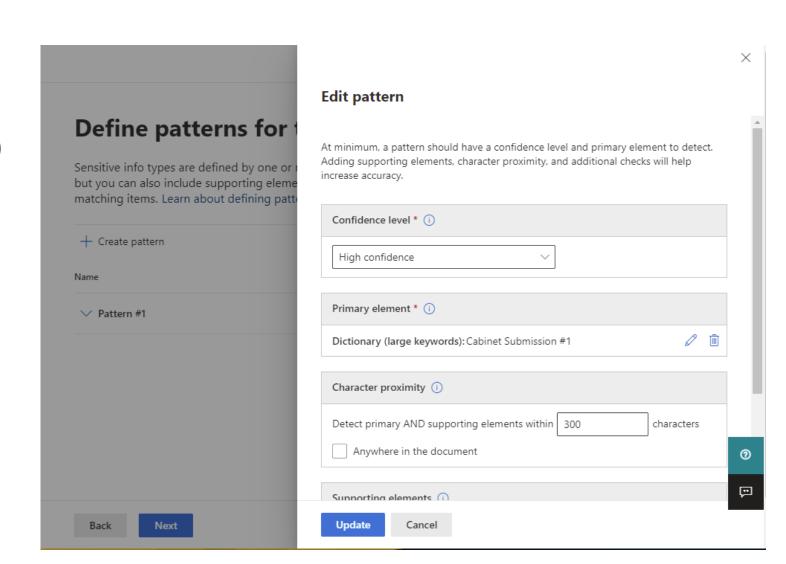
Cancel

Sensitive Info Types (SIT's):

Creating a SIT requires
 Define the pattern –
 here's where we get
 serious



- Creating a SIT requires =
- + Edit pattern
- Set for low confidence (for now)
- + Can be tweaked later...

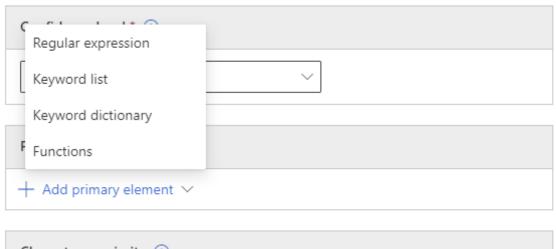


Sensitive Info Types (SIT's):

- Creating a SIT requires =
- + New/Edit pattern
- + What to choose here?

New pattern

At minimum, a pattern should have a confidence level and primary element to detect. Adding supporting elements, character proximity, and additional checks will help increase accuracy.



Character proximity ()		
Detect primary AND supporting elements within	300	characters
Anywhere in the document		



Sensitive Info Types (SIT's):

- + SIT can be based on the following:
 - + Regular Expressions (not covered here see speaker notes)
 - **Keyword Lists** use by default unless you have phrases over 50 characters
 - Use keyword lists when you want have few keywords (limit is 2K keywords) and each keyword is less than 50 characters.
 - + Keyword lists have granular features like choosing to do a string match vs full word match, case sensitive vs case insensitive etc.

Keyword Dictionary

- Use keyword dictionaries when you have thousands of phrases some of which might be over 50 characters long.
- + Functions (not covered here see speaker notes)

Sensitive Info Types (SIT's):

- Creating a SIT
- + Add a Keyword List:

Add a keyword list

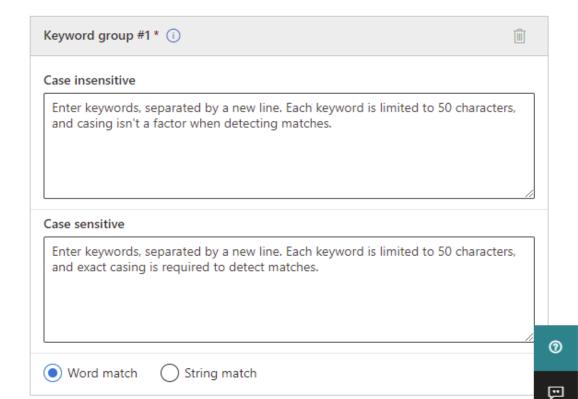
Keyword lists identify the words and phrases you want this info type to detect. For example, the keyword list to identify Netherlands VAT numbers is 'VAT number, vat no, vat number, VAT#'. Learn how to create keyword lists

E Choose from existing keyword lists

Cancel

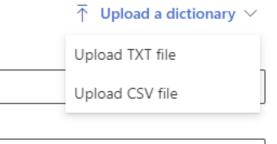
ID * (i)

Enter the keyword list name



Sensitive Info Types (SIT's):

- Creating a SIT
- + Add a Keyword Dictionary:
- + HINT:
- Take example doc & save as text using Unicode – this can now be used as an upload
- Remove all words that may cause noise to reduce false positives



'd dictionary will be saved for use in

Add keyword dictionary

Unlike keyword lists (which are limited in size) keyword dictionaries provide easier management of keywords and at a much larger scale. Learn how to create keyword dictionaries

E Choose from existing dictionaries

→ Upload a dictionary ∨

Name * (i)

<Name of Dictionary>

Keywords * (i)

APPENDIX A

OFFICE USE ONLY

CONFIDENTIAL

CABINET SUMMARY SHEET

SUBMISSION NO.

DATE RECEIVED

TITLE OF CABINET MINUTE

MINISTER'S NAME, TITLE AND RELEVANT PORTFOLIO

IMPLEMENTING AGENCY

PURPOSE (objective of proposal)

COSTING/FINANCIAL IMPLICATIONS

Is proposal covered by existing/approved forward estimates?

IF NO, DOES PROPOSAL HAVE AN ADDITIONAL IMPACT ON:

Expense Limit

Net Operating Balance

Net Debt

FTE Increases

IS PROPOSAL TO BE FUNDED (FULLY OR PARTIALLY) VIA:

Re-prioritisation of savings

Reduction in cash balance

Royalties for Regions

Increase in appropriation

Increase in retained revenue

۳

0

Done

Cancel

Sensitive Info Types (SIT's):

- Creating a SIT
- + Add a Keyword Dictionary:
- Compare the new Keyword Dictionary to the original Doc and notice the intersection?

		אוו בוזעוא א	
	OFFICE USE ONLY		
CONFIDENTIAL	SUBMISSION NO.	DATE RECEIVED	
CABINET SUMMARY SHEET			
TITLE OF CABINET MINUTE			
MINISTER'S NAME, TITLE AND RELEVANT PORTFOLIO			
IMPLEMENTING AGENCY			
PURPOSE (objective of proposal)			
COSTING/FINANCIAL IMPLICATIONS Is proposal covered by existing/approved forward estimates?	□ Yes	□ No	
IF NO, DOES PROPOSAL HAVE AN ADDITIONAL IMPACT ON:	t Debt	FTE Increases	
Yes No Yes No	Yes No	Yes No	
IS PROPOSAL TO BE FUNDED (FULLY OR PARTIALLY) VIA: Re-prioritisation of savings Reduction in ca	sh balance		
Royalties for Regions Increase in appropriation Increase in reta	ined revenue		

Add keyword dictionary

Unlike keyword lists (which are limited in size) keyword dictionaries provide easier management of keywords and at a much larger scale. Learn how to create keyword dictionaries

E Choose from existing dictionaries

↑ Upload a dictionary ∨

Name * (i)

<Name of Dictionary>

Keywords * (i)

ΔΡΡΕΝΙΠΙΥ Δ

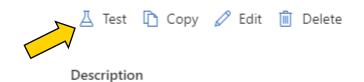
APPENDIX A OFFICE USE ONLY CONFIDENTIAL CABINET SUMMARY SHEET SUBMISSION NO. DATE RECEIVED TITLE OF CABINET MINUTE MINISTER'S NAME, TITLE AND RELEVANT PORTFOLIO IMPLEMENTING AGENCY PURPOSE (objective of proposal) COSTING/FINANCIAL IMPLICATIONS Is proposal covered by existing/approved forward estimates? IF NO, DOES PROPOSAL HAVE AN ADDITIONAL IMPACT ON: Expense Limit Net Operating Balance Net Debt FTE Increases IS PROPOSAL TO BE FUNDED (FULLY OR PARTIALLY) VIA: Re-prioritisation of savings Reduction in cash balance Royalties for Regions 0 Increase in appropriation Increase in retained revenue

Done Cancel

Sensitive Info Types (SIT's):

- + Creating a SIT
- + Add a Keyword Dictionary:
- Once created use the test button
- + Upload the original docx
- Then start modifying docx and retesting
- + Tweak and retest
- Retest on true and false positive examples

TEST - Cabinet Submission #1



Match results

We have detected the following in Example_APPENDIX A.docx

1. TEST - Cabinet Submission #1

Low - 45 matches	
Matches	Supporting elements
(Regulatory proposals with Economically Significant	-
(If yes, complete Regulatory Impact Assessment secti	-
MINISTER'S NAME, TITLE AND RELEVANT PORTFOLIO	-
Re-prioritisation of savings	-
PURPOSE (objective of proposal)	-
(to be detailed in Consultation section of cabinet mi	-
Consultation	-
Regulatory Impact Assessment	-
TIMING OF ANNOUNCEMENT	-

Sensitive Info Types (SIT's):

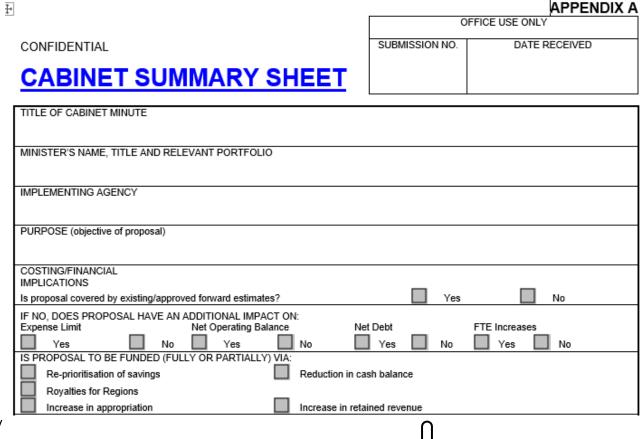
- Always evaluate ALL doc's that are
 mandated to be in the solution
 - Having said that that's for a Prod deployment
 - For a PoC best to be agile and nimble - so it comes down to picking a "good" demo doc
 - The more precise & focused the better

Ţ-•		APPENDIX A
	О	PFFICE USE ONLY
CONFIDENTIAL	SUBMISSION NO.	DATE RECEIVED
CABINET SUMMARY SHEET		
TITLE OF CABINET MINUTE		
MINISTER'S NAME, TITLE AND RELEVANT PORTFOLIO		
IMPLEMENTING AGENCY		
PURPOSE (objective of proposal)		
COSTING/FINANCIAL IMPLICATIONS		
Is proposal covered by existing/approved forward estimates?	Yes	☐ No
IF NO, DOES PROPOSAL HAVE AN ADDITIONAL IMPACT ON: Expense Limit Net Operating Balance Ne	t Debt	FTE Increases
Yes No Yes No	Yes No	Yes No
IS PROPOSAL TO BE FUNDED (FULLY OR PARTIALLY) VIA: Re-prioritisation of savings Reduction in ca	ish balance	
Royalties for Regions Increase in appropriation Increase in reta	ined revenue	

Something like this

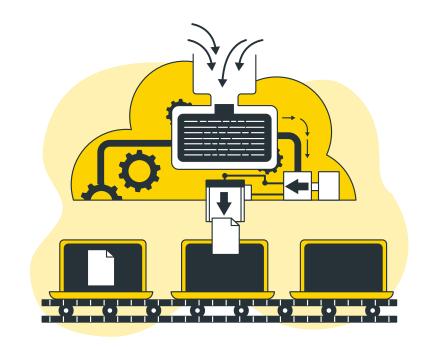
Sensitive Info Types (SIT's):

- Always evaluate ALL doc's that are
 mandated to be in the solution
- Creating a SIT can be 3 different Patterns - this is only needed if there are a "number" of User Stories and Doc's of differing degrees
 - High/Med/Low choosing between these comes down to the Org's risk appetite + details above
 - Different patterns may alert at different levels in business workflow
 - For a PoC, this is likely not needed so use the KISS principal



Something like this

- When creating the SIT be aware that the "Test" should be a considered a "True Funnel" of what will match
 - If you create a High Confidence SIT and then "Test" = you will get matches for Low/Med/High (It's a True Funnel)
 - If you create a Low Confidence SIT and then "Test" = you will get matches for Low only (It's a True Funnel)

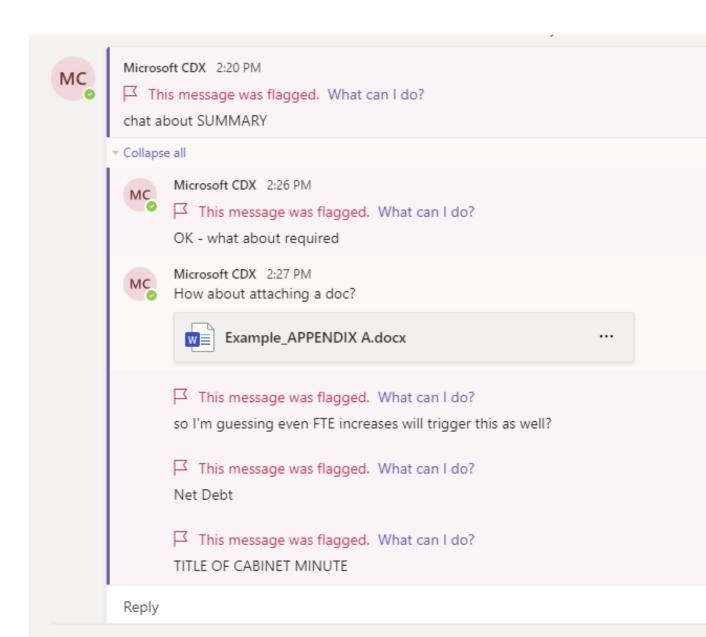


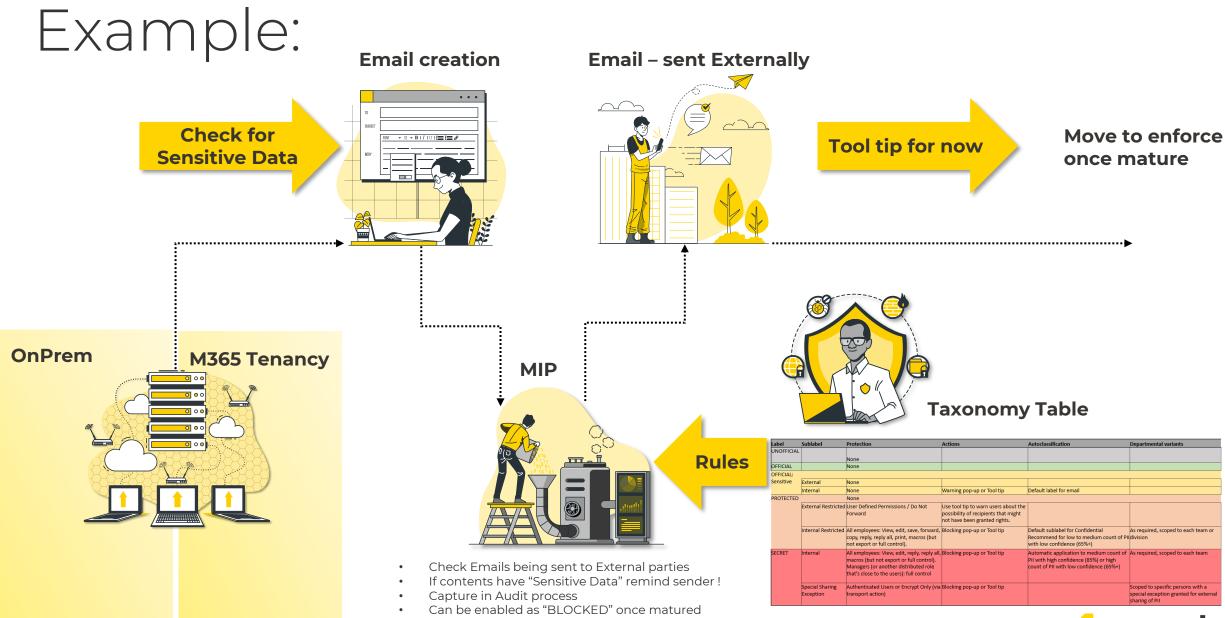


- → So when creating the SIT make sure you don't tweak it too far that you accidently eliminate any True Positives
 - + Be mindful that a SIT is **simply the first cut** the next step is to use that SIT from either a DLP Policy or a policy to apply a Sensitivity or Retention Label
 - Both of these "Polices" have filters that can be used for filtering down to the "stuff" that is relevant
 - + So make sure you haven't gone too far in the SIT, you may be missing TP's?
 - + That's not good so you may need to go back to the drawing board...
 - → But at the same token you don't want false positives to be a couple of orders of magnitude this needs balance



- Further testing enable policy in Teams
- Now use Teams (in browser) and start referencing the keywords from that Keyword Dictionary example created earlier in a random chat
- Here we can test in almost real-time the responses











'empired_

Understanding sensitivity labels

Customizable

Persists as container metadata or file metadata

- Readable by other systems

Determines DLP policy based on labels



Manual or Automated Labels

Apply to content or containers

Label data at rest, data in use, or data in transit

Enable protection actions based on labels

Seamless end user experience across productivity applications

Extensible to partner solutions



Label	Sublabel	Protection	Actions	Autoclassification	Departmental variants
UNOFFICIAL					
		None			
OFFICIAL		None			
OFFICIAL:					
Sensitive	External	None			
	Internal	None	Warning pop-up or Tool tip	Default label for email	
PROTECTED		None			
			Use tool tip to warn users about the possibility of recipients that might not have been granted rights.		
		All employees: View, edit, save, forward, copy, reply, reply all, print, macros (but not export or full control).		Default sublabel for Confidential Recommend for low to medium count of PII with low confidence (65%+)	As required, scoped to each team or Idivision
SECRET		All employees: View, edit, reply, reply all, macros (but not export or full control). Managers (or another distributed role that's close to the users): full control		Automatic application to medium count of PII with high confidence (85%) or high count of PII with low confidence (65%+)	As required, scoped to each team
		Authenticated Users or Encrypt Only (via transport action)	Blocking pop-up or Tool tip		Scoped to specific persons with a special exception granted for external sharing of PII

Polated DLD rules

Trigger	Example	DLP actions	Exception	Monitoring
Low count PII	1-2 Credit Card #s, SSNs, etc.	Block external sharing	If classified as non-business, Confidential External or HC Special Sharing Exception	Alert on repeat actions in a short period of time.
Medium count PII	3-9 Credit Card #s, SSNs, etc.	Block external sharing, Cloud Egress, copy to removable drive	If classified as Confidential External or HC Special Sharing Exception	Alert on repeat actions over time.
High count PII	10+ Credit Card #s, SSNs, etc.	Block external sharing, Cloud Egress, copy to removable drive	If classified as HC Special Sharing Exception (and/or sent to special allowed parties by specific allowed users)	Report on each individual action.

Information **Protection &** Governance

Protect and govern data – wherever it lives

Understand your data landscape and identify important data across your hybrid environment



Apply flexible

protection actions

access restrictions

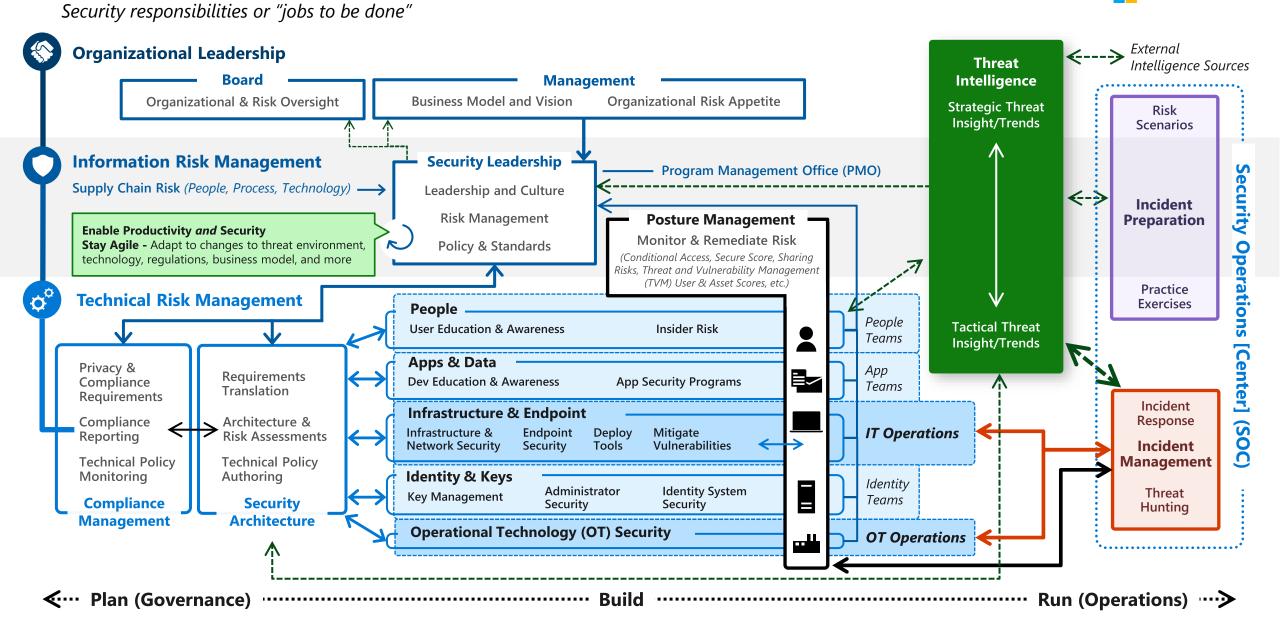
and visual markings

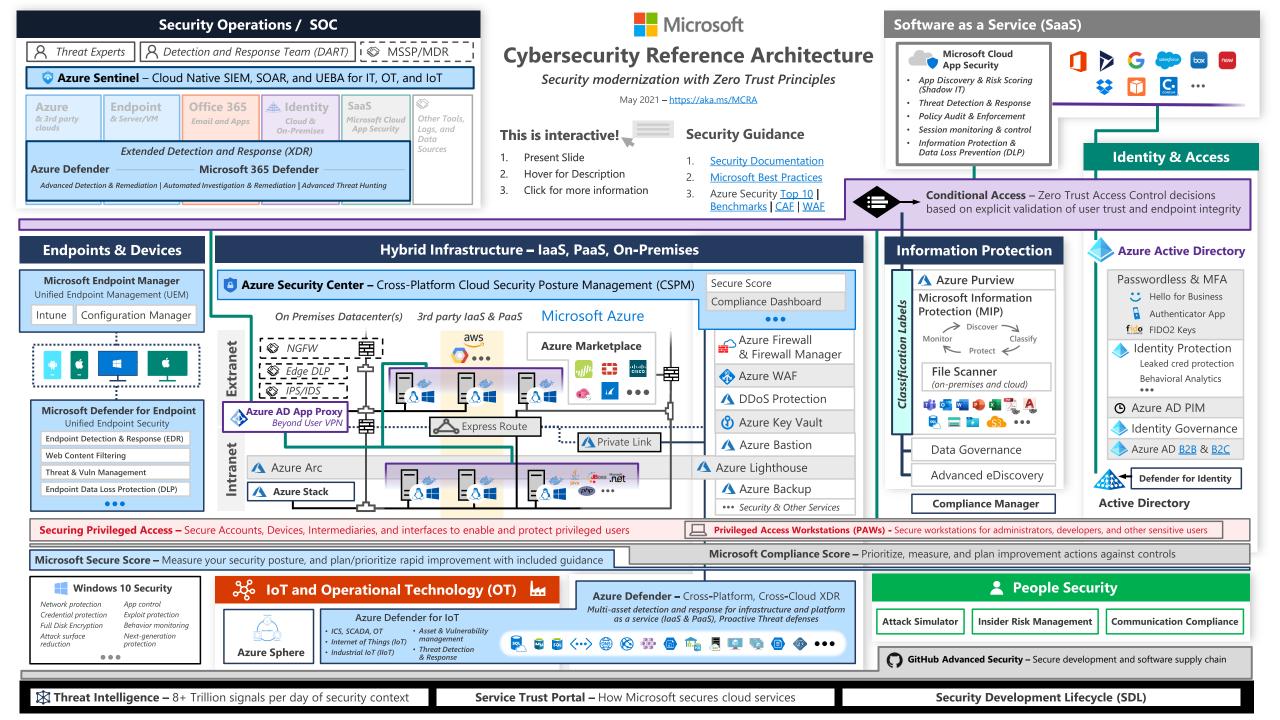
Automatically retain, delete, and store data and records in compliant manner

Powered by an intelligent platform

Unified approach to automatic data classification, policy management, analytics and APIs







Cross-cloud and cross-platform



Comprehensive Security, Compliance and Identity capabilities that integrate with your existing solutions

May 2021 - https://aka.ms/MCRA

Industry Partnerships

NIST / CIS / The Open Group / Others

Microsoft Intelligent Security Association

Solution Integration and MDR/MSSP Partners

CERTs / ISACs / Others

Law Enforcement



Microsoft Security, Compliance, and Identity Capabilities

Threat Intelligence – 8+ Trillion signals per day of security context

Access Control Identity and Network **Modern Security Operations**

Rapid Resolution with XDR, SIEM, SOAR, UEBA and more

Asset Protection

Information Protection and App Security / DevSecOps

Technical Governance

Risk Visibility, Scoring, and Policy Enforcement

People Security – User Education/Empowerment and Insider Threats





















































Software as a Service (SaaS)

















Hybrid Infrastructure – IaaS, PaaS, On-Premises







IoT Devices







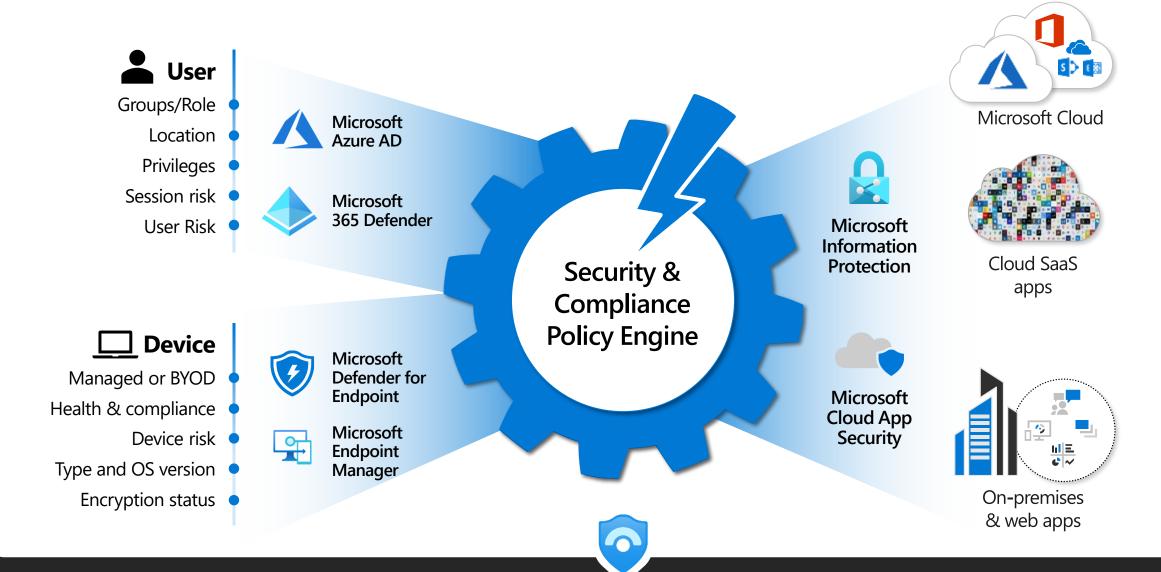


Operational Technology (OT)

Security Operations [Center] (SOC) – Reduce attacker time/opportunity to impact business

Protect assets anywhere with Zero Trust

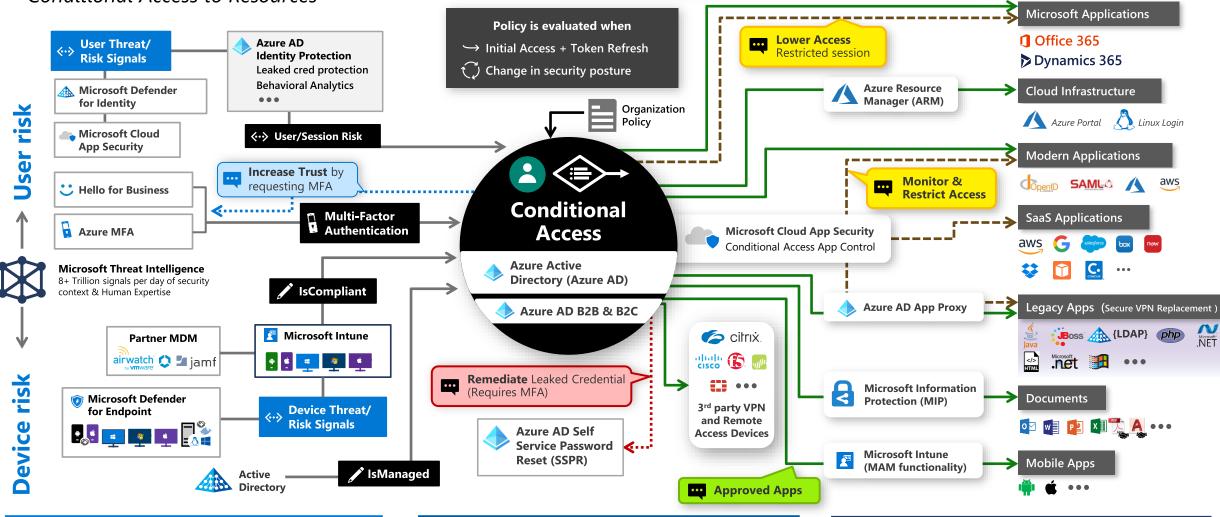
Verify explicitly | Use least-privileged access | Assume breach

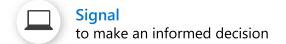


Zero Trust User Access

Legend Full access --- Limited access
Risk Mitigation Remediation Path

Conditional Access to Resources









Microsoft

May 2021 - https://aka.ms/MCRA