RSA Conference2016

San Francisco | February 29 – March 4 | Moscone Center

Connect to Protect

SESSION ID: PDAC-R03

Data Classification: Reclaiming Infosec's Redheaded Stepchild

Yuval Eldar

Founder, Secure Islands

@SecureIslands

Microsoft | secure islands

Reclaiming Infosec's Redheaded Stepchild

Why was classification neglected until now?

- Does your organization have data classification policies?
- What percentage of your data is being classified?

The sad reality…

55%

of IT professionals say data classification is too complex to plan, manage and deploy

63%

of IT professionals are not certain that their company's classification scheme is aligned with how data is created, used and shared

88%

of IT professionals say they ignored or circumvented data classification policies

* Based on a survey of 100 IT professionals conducted by Secure Islands, Nov. 2015

Why is classification so critical … Now more than ever

Information security starts with

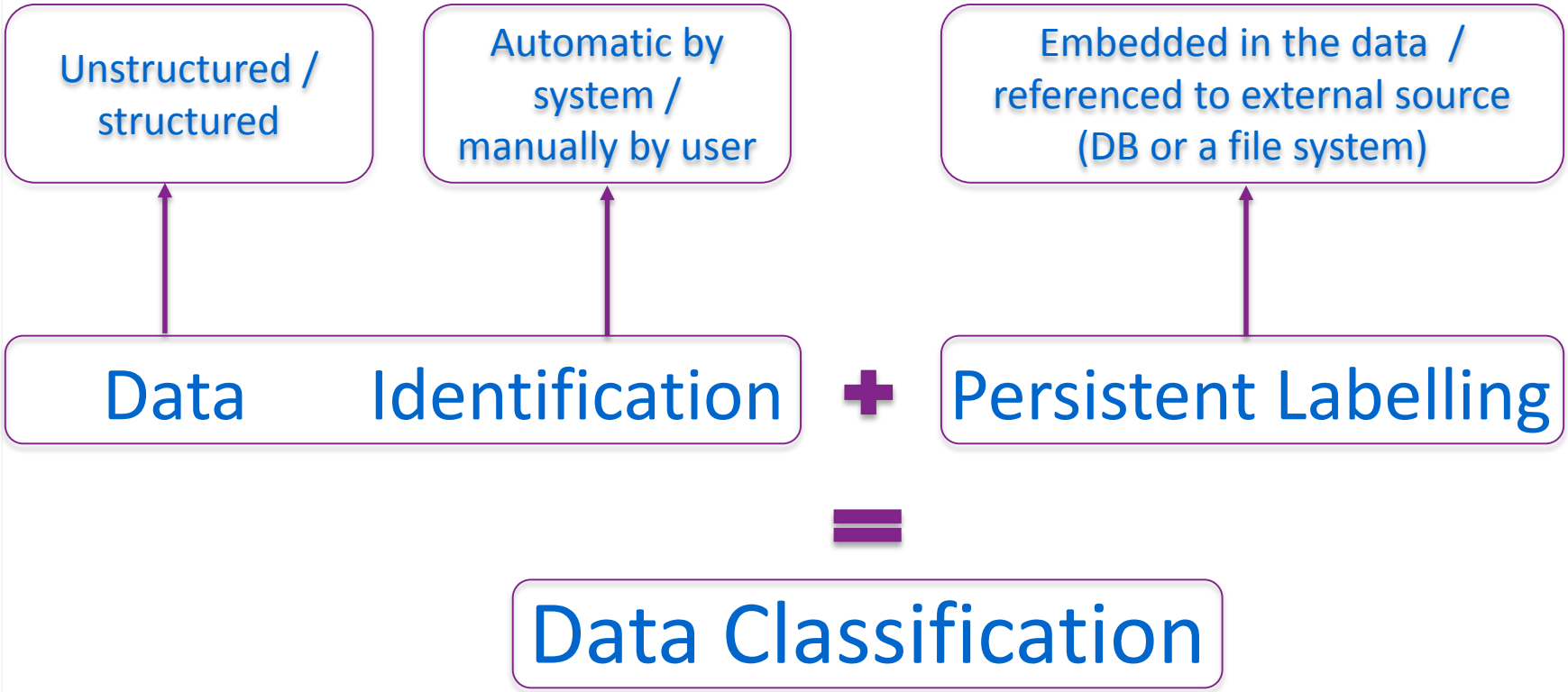
CLASSIFICATION

The CISO dilemma…

"I cannot start my <fill in the blank> project before I know how to identify my (sensitive) data…"

- A) IRM
- B) DLP
- C) Access controls
- D) Mail encryption
- E) Moving to the Cloud
- F) Data retention
- G) All of the above

What is an effective data classification model?

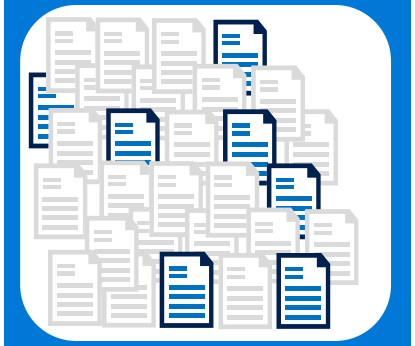


It’s not “We Should”…

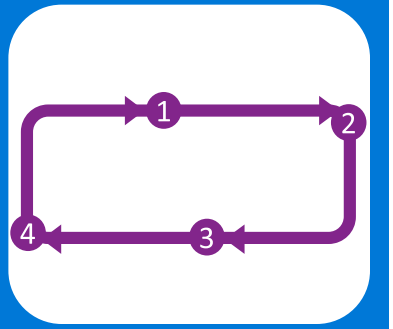
It’s “We Can”!

The 4 basic steps for implementing data classification

Define what to classify



Decide in which stage to classify



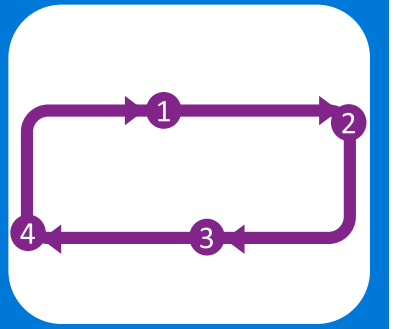
Select the method of classification (manual/automatic)

Define and apply the data class labels

Step 1: Define what to classify

Define what to classify

Decide in which stage to classify



Select the method of classification (manual/automatic)

Define and apply the data class labels

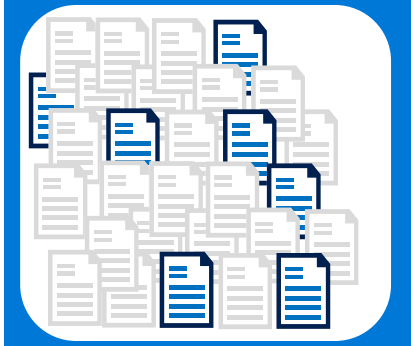
Deciding what to classify

Not all data was created equal!

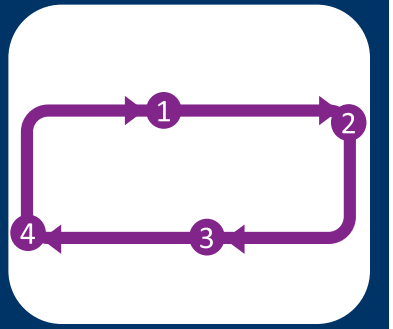
- Don’t try to classify all your data
- Concentrate on your high business impact first
- Remember that this is an ongoing, iterative process

Step 2: Decide in which stage to classify

Define what to classify



Decide in which stage to classify



Select the method of classification (manual/automatic)

Define and apply the data class labels

In which stage to classify?

22004	Baits, Donald	Public Relations	PR
33012	Brenin, Kenny	Tech Support	Tech
33005	Burr, Kyle	Tech Support	Tech
33013	Cade, Arnie	Tech Support	Office Clerk
11004	Cann, Bud	Administration	Tech
22003	Carwin, Eric	Public Relations	PR
33009	Darrok, Jimmy	Tech Support	Tech
22009	Doe, John	Sales	Sales Rep
33004	Fales, Hommer	Shipping	Shipping
33001	Fane, Lisa	Sales	Sales Rep
33008	Fischer, Jeff	Tech Support	Tech
33010	Hacker, Helga G.	Tech Support	Tech
11003	Hallam, Clark	Administration	Office Clerk
22001	Keen, Luke	Graphics	Designer
33002	Kemp, Milhouse	Sales	Sales Rep
33006	Lucius, Eddie	Tech Support	Tech
11002	Napier, Marge	Administration	Office Clerk
11001	Nevin, Harry	Administration	Office Clerk
22008	Quntin, Ralph	Sales	Sales Rep
33003	Ramsden, Clark	Shipping	Shipping
33007	Rankin, Bart	Tech Support	Tech
22002	Sabina, Mick	Graphics	Designer
22007	Saxton, Catherine	Public Relations	PR
22006	Smith, Jane	Public Relations	PR
11009	Tabor, Fred	Graphics	Designer
11006	Taylor, Max	Administration	Office Clerk
33011	Urran, Stan	Tech Support	Tech
11005	Ursula, Saul	Administration	Office Clerk
11008	Yager, Rodney	Graphics	Designer
22005	Yeo, Pancho	Public Relations	PR

Can you guess what this list represents??

In which stage to classify?

22004	Baits, Donald	Public Relations	PR
33012	Brenin, Kenny	Tech Support	Tech
33005	Burr, Kyle	Tech Support	Tech
33013	Cade, Arnie	Tech Support	Office Clerk
11004	Cann, Bud	Administration	
22003	Carwin, Eric	Public Relations	
33009	Darrok, Jimmy	Tech Support	
22009	Doe, John	Sales	
33004	Fales, Hommer	Shipping	
33001	Fane, Lisa	Sales	
33008	Fischer, Jeff	Tech Support	
33010	Hacker, Helga G.	Tech Support	
11003	Hallam, Clark	Administration	
22001	Keen, Luke	Graphics	
33002	Kemp, Milhouse	Sales	
33006	Lucius, Eddie	Tech Support	
11002	Napier, Marge	Administration	
11001	Nevin, Harry	Administration	
22008	Quntin, Ralph	Sales	
33003	Ramsden, Clark	Shipping	
33007	Rankin, Bart	Tech Support	
22002	Sabina, Mick	Graphics	
22007	Saxton, Catherine	Public Relations	
22006	Smith, Jane	Public Relations	
11009	Tabor, Fred	Graphics	
11006	Taylor, Max	Administration	
33011	Urran, Stan	Tech Support	
11005	Ursula, Saul	Administration	
11008	Yager, Rodney	Graphics	
22005	Yeo, Pancho	Public Relations	

Now, is it clearer?

Classify as close to the source as possible

The first step in identifying sensitive data is to examine its source at creation

Classification based on the **context** of the **source** results in **accuracy**

The data owner is **accountable**

Starting **at birth** allows to apply protection as early in the lifecycle as possible and covers the **entire info lifecycle**

How to accomplish this step?

From your high business impact data:

Identify sources

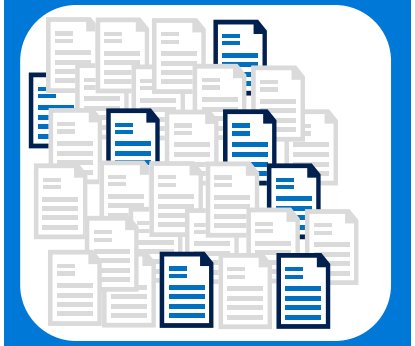
- Applications
- File servers
- Databases
- Repositories

Valuable info can be deduced from other initiatives like:

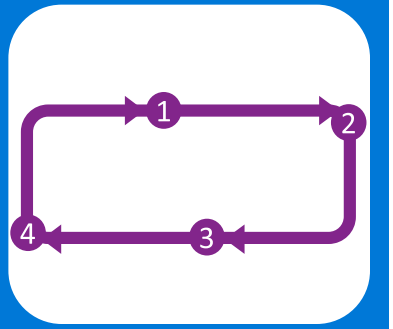
- Audit reviews
- Risk analysis reports
- Etc.

Step 3: Select the method of classification

Define what to classify



Decide in which stage to classify



Select the method of classification (manual/automatic)

Define and apply the data class labels

What method to use?

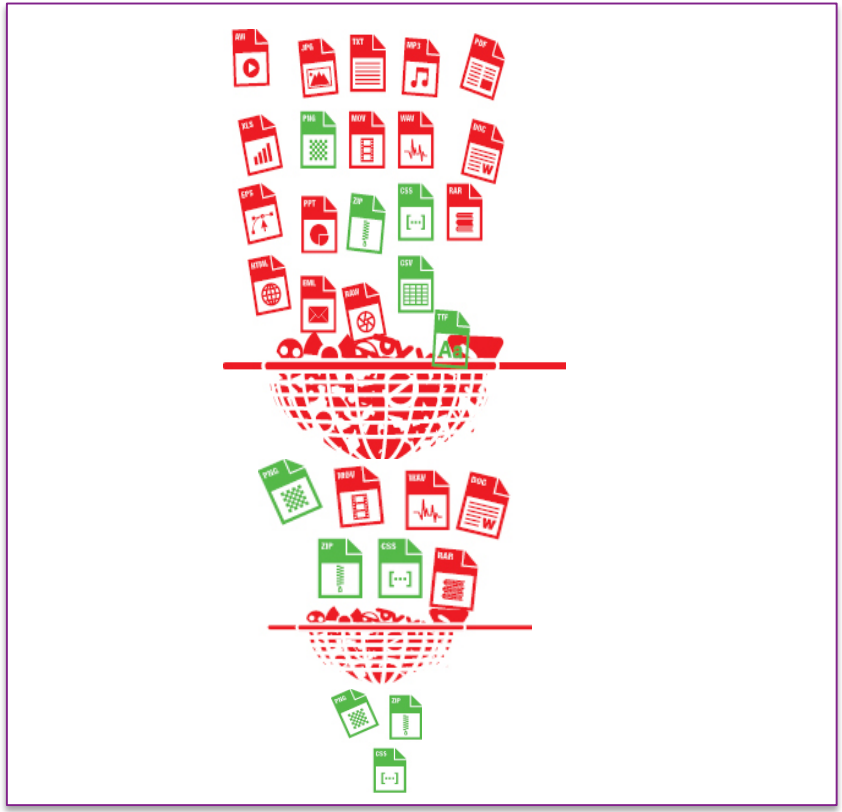
The aspiration -> Minimize the friction with the end user

Minimizing friction with the user

Automatic seamless classification

System recommendation

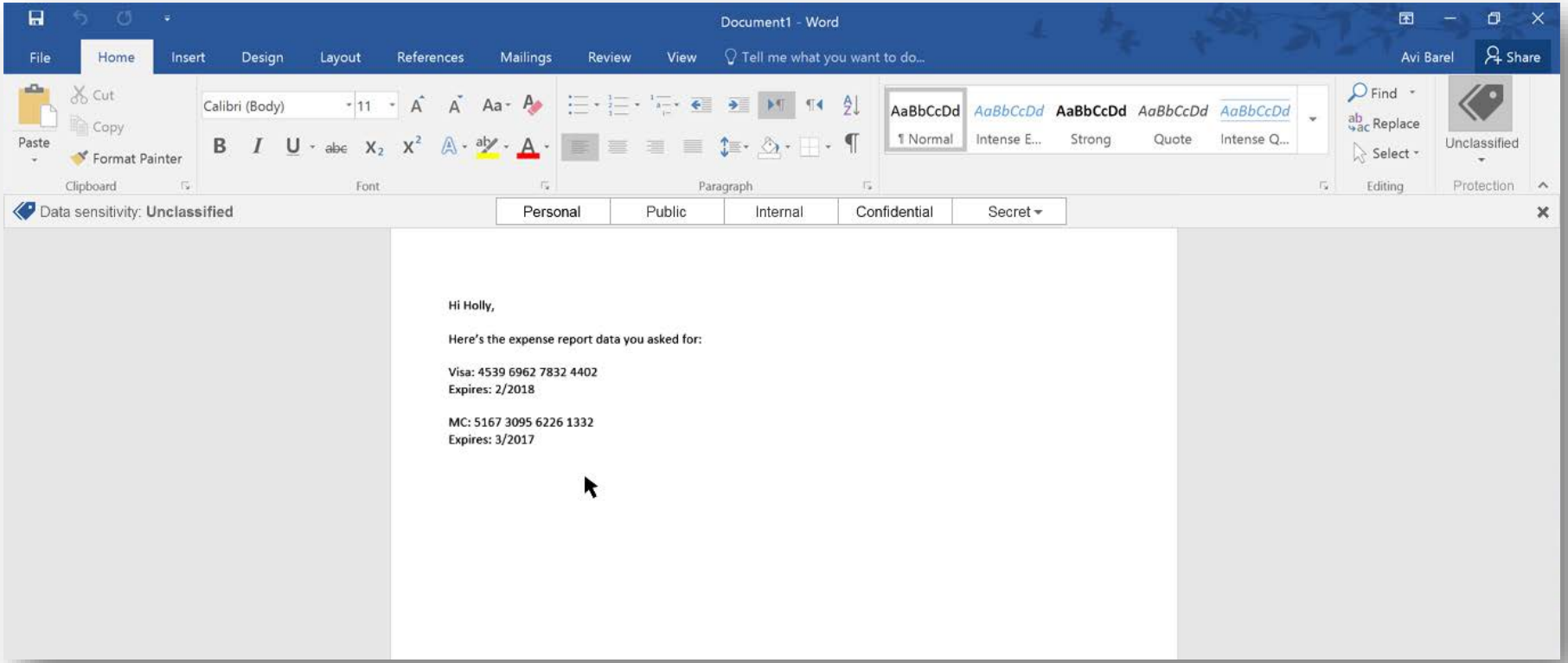
User driven

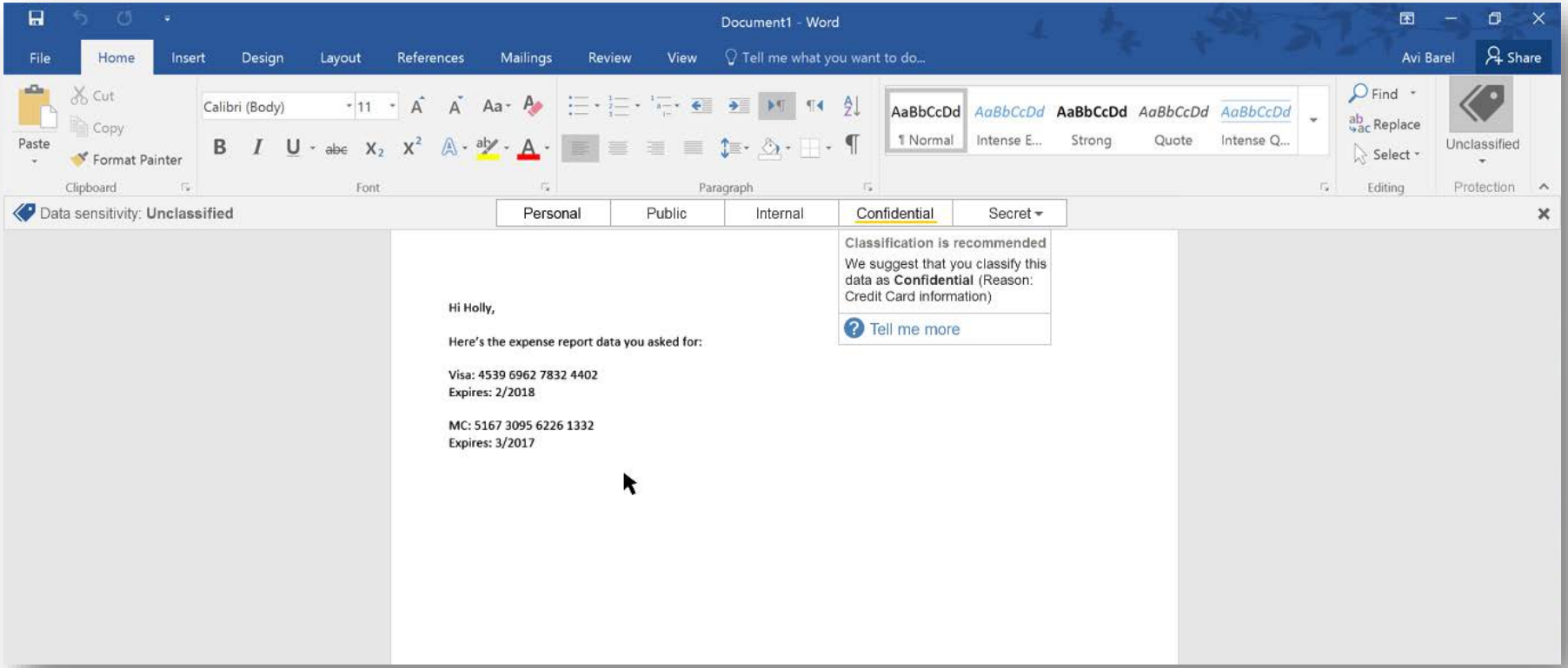


Methods of Information Classification

User driven classification

- ☑ May classify a document in an accurate way when working on it
- ☒ Classification may not be predicted across the org (it is manual process after all)
- ☒ Users forget to classify and may object to the process
- ☒ Users' frustration and lack of effectiveness over time


Document1 - Word
File
Home
Insert
Design
Layout
References
Mailings
Review
View
Tell me what you want to do...
Avi Barel
Share
Cut
Copy
Format Painter
Paste
Clipboard
Calibri (Body)
11
Font
AaBbCcDd
¶ Normal
AaBbCcDd
Intense E...
AaBbCcDd
Strong
AaBbCcDd
Quote
AaBbCcDd
Intense Q...
Paragraph
Find
Replace
Select
Editing
Unclassified
Protection
Data sensitivity: Unclassified
Personal
Public
Internal
Confidential
Secret
Hi Holly,
Here's the expense report data you asked for:
Visa: 4539 6962 7832 4402
Expires: 2/2018
MC: 5167 3095 6226 1332
Expires: 3/2017


Document1 - Word
File
Home
Insert
Design
Layout
References
Mailings
Review
View
Tell me what you want to do...
Avi Barel
Share
Cut
Copy
Format Painter
Paste
Clipboard
Calibri (Body)
11
Font
AaBbCcDd
¶ Normal
AaBbCcDd
Intense E...
AaBbCcDd
Strong
AaBbCcDd
Quote
AaBbCcDd
Intense Q...
Paragraph
Find
Replace
Select
Editing
Unclassified
Protection
Data sensitivity: Unclassified
Personal
Public
Internal
Confidential
Secret
Classification is recommended
We suggest that you classify this data as Confidential (Reason: Credit Card information)
Tell me more
Hi Holly,
Here's the expense report data you asked for:
Visa: 4539 6962 7832 4402
Expires: 2/2018
MC: 5167 3095 6226 1332
Expires: 3/2017

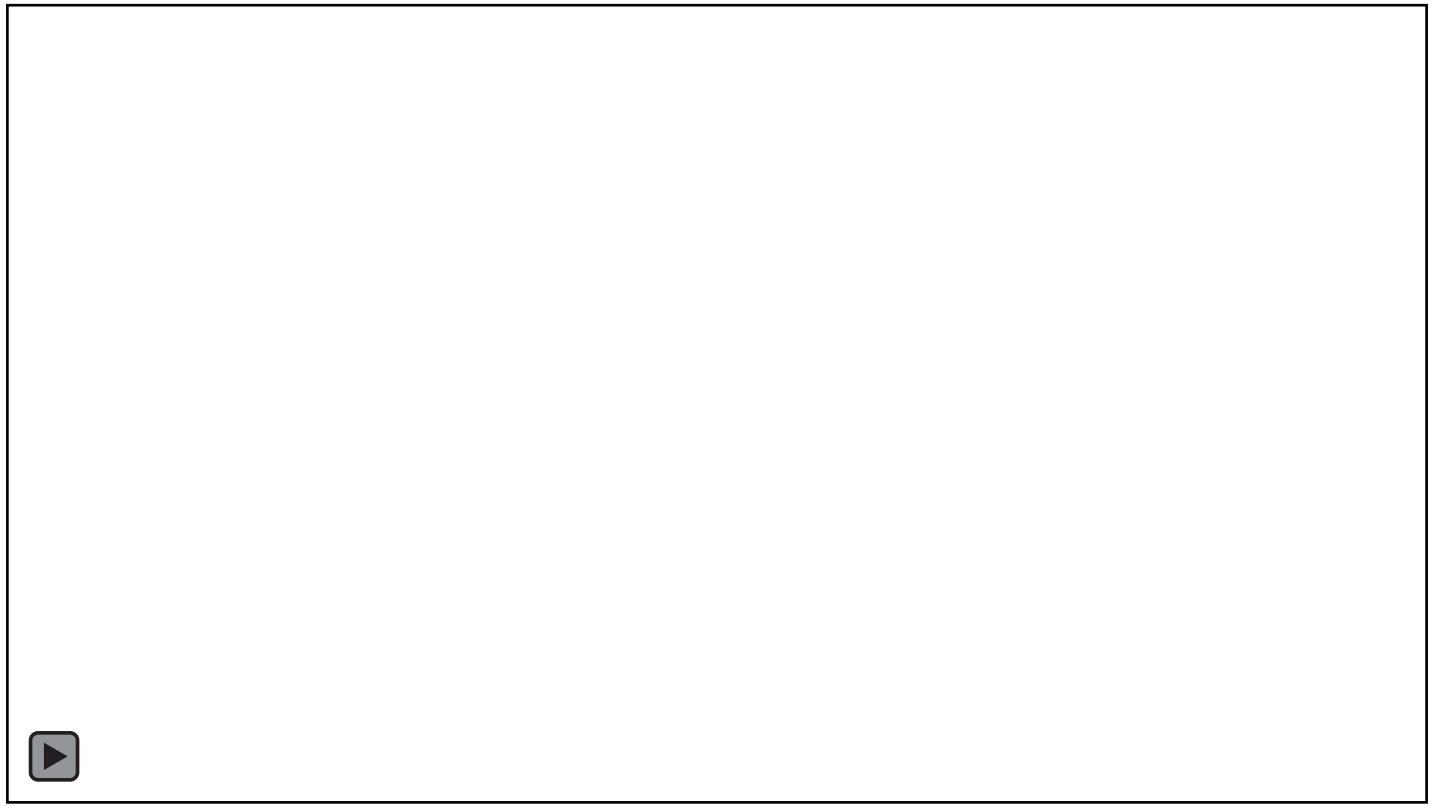
Methods of Information Classification

User driven classification

Source based automatic classification

- ☑ Classification at the source – where information is created
- ☑ Classify data created by any source at the business
- ☑ 100% accurate. Always Predictive
- ☒ Requires pre-data mapping -> admin should define policies/rules

Automatic classification demo



Data classification examples

Intercept files at the source, upon creation

Financial advisor +

Financial report from SAP

Customers' ID patterns +

salesforce

Salesforce report

Files copied to the M&A folder in SharePoint Online

Methods of information classification

New concept:
“Crowdsourcing Classification”

Automatic classification

User driven classification

Crowdsourcing-based classification

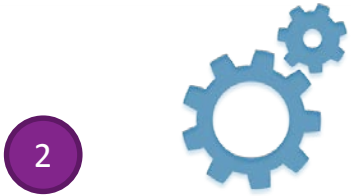
AUTOMATIC POLICIES

CROWD GENERATED

How does it work?

User classifies a document/file manually

USER
CONTENT
CONTEXT
RESULT



The system based on Machine Learning engine learns the classification env. (content, context, and classification)

Additional users classify similar data in similar way

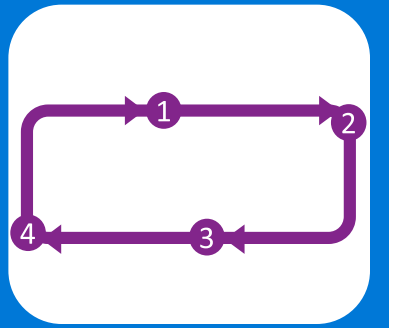
USER
CONTENT
CONTEXT
RESULT

Generate automatic Classification for this data type

Step 4: Define and apply the data class labels

Define what to classify

Decide in which stage to classify



Select the method of classification (manual/automatic)

Define and apply the data class labels

Which data-class labels to apply?

* For “DLP” use-cases, sensitivity levels is enough (public, internal, confidential, secret)
  For SoD/Internal compartmentalization, multi dimensional labels should be needed

Define the required classification labels

- What is the **minimal** set of Classification labels necessary to convey the protection?
- Distinguish between different types of Classification records:
- List the levels according to the order of their sensitivity
- Consider one record for each protection policy

- **In most cases it is possible, and recommended, to use only sensitivity levels labels!**

Sensitivity Levels:	Classification Subjects:	Classification Flags:
Public	HR Info	Cross Border
Internal	CID Info	Country Segregated
Confidential	Finance Info	External Comm.
Secret	Others?	Waiver?

Define the required protection policy

What classification labels are required to support your protection needs?

- Build a Classification matrix with suitable protection policies

Example A

Classification Level	Classification Subject	Classification Flag	Protection policy
Public		-	None
Internal			All Employees
Confidential		-	All FTE employees
Secret	Finance Info	-	Finance Group

Define the required protection policy

Example B

Classification Level	Classification Subject	Classification Flag	Protection Policy
Internal		-	All Employees
Confidential	CID Info	Country X	Employees in Country X only
Confidential	Finance Info	-	Finance & Management only
Public	Finance Info	-	None

Some tips for effective information classification

Tips for effective information classification

1. Choose a solution that allows both manual AND automatic classification
2. Make sure to choose a solution that covers all data sources (including LoB apps) and is not focused on MS-Office alone
3. Use a classification scheme that leverages and enhances existing tools such as DLP, archiving, e-discovery and more
4. Use persistent labelling that follows the data wherever it goes and throughout its entire lifecycle (be platform agnostic)

Apply

Apply What You Have Learned Today

- Next week you should:
  - Identify your high business impact data within your organization
- In the first 3 months following this presentation you should:
  - Understand from what sources this data is being generated/accessed
  - Define a classification scheme which correlates your protection policies
  - Review classification systems (also inquiry analyst firms in this field)
- Within 6 months you should:
  - PoC-ing/pilot-ing a security system which can “intercept” different sources with minimum friction with the end user

Questions?

Data Classification: Reclaiming Infosec's Redheaded Stepchild

Thank You

Yuval Eldar
Founder, Secure Islands