

# MODERNIZE MIGRATION AND INFORMATION CLASSIFICATION WITH INFOTECHTION

Natalie Noonan, Global Black Belt, Microsoft  
Brian Tuemmler, Founding Partner, Infotechtion  
Vivek Bhatt, Founder and CTO, Infotechtion



 **Infotechtion**



## Locations

- **United States:** Infotechion LLC  
Company Number 202008610635
- **UK:** Infotechion Limited  
Company Number 12048948
- **Netherlands:** Infotechion BV  
Company Number 80933033
- **Norway:** Infotechion AS  
Company Number 914124972

Infotechion

## Corporate Background

## Company Overview

- Microsoft partners specializes in improving and automating information governance processes: retention, protection & governance, leveraging Microsoft Cloud
- Microsoft GOLD Partner
- Microsoft cloud services, Project Cortex and Compliance partner
- IBM services partner for Information governance consulting

## Achievements

- **2019:** Company and clients consolidation as part of formation
- **2020:** First Microsoft reference to have delivered Microsoft Office 365 based records management at enterprise scale
- **2020:** 0 Debt, 100% growth in staff and on track to deliver forecasted sales and profit margins
- **2020:** Registered supplier on UK Government cloud framework

# OUR SERVICES

Value from  
Information

Minimize  
Compliance risks

Create new  
Opportunities

Optimize Adoption  
and Operational Costs

## INFORMATION GOVERNANCE STRATEGY

Vision

Use Cases

Success  
Factors

KPIs

Roadmap

## INFORMATION GOVERNANCE ARCHITECTURE

Taxonomy

Metadata

Security

Search

Records

## INFORMATION GOVERNANCE SOLUTIONS

Compliance  
Assessments

Governance  
Accelerators

AI  
Automation

Learning  
Solutions

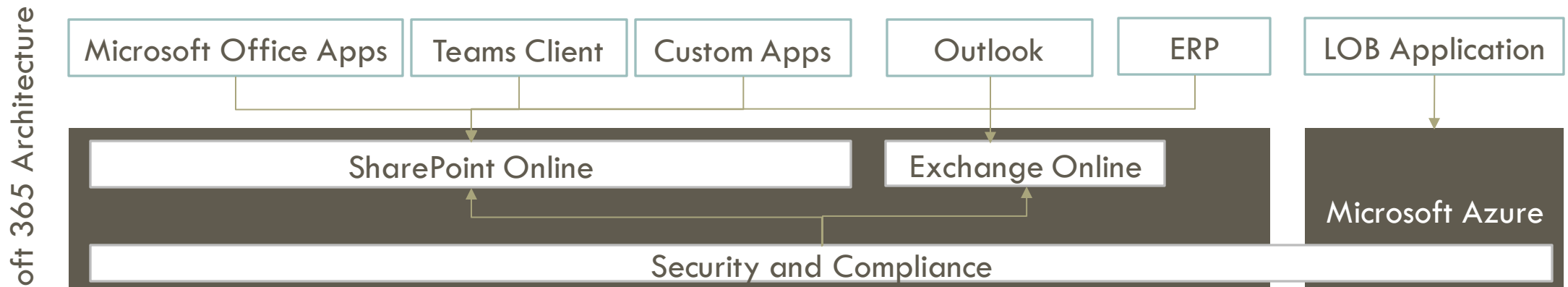
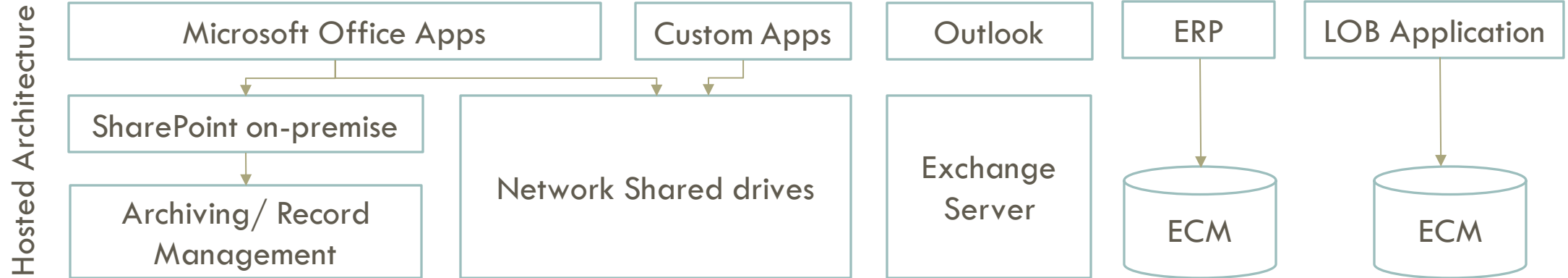
Migration  
Solutions

Information Management Practices and Behaviours

*"I consider Infotechtion one of the leading experts in Microsoft information governance. Their staff work closely with our enterprise customers to maximize customer investment in Microsoft products especially Microsoft information protection and governance to enable increased levels of compliance for customer information in Microsoft 365."*

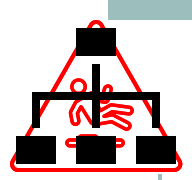
Principal Engineering Manager, S+C Engineering, Microsoft.

# THE CONTEXT: ARCHITECTURAL TRANSITION TO CLOUD

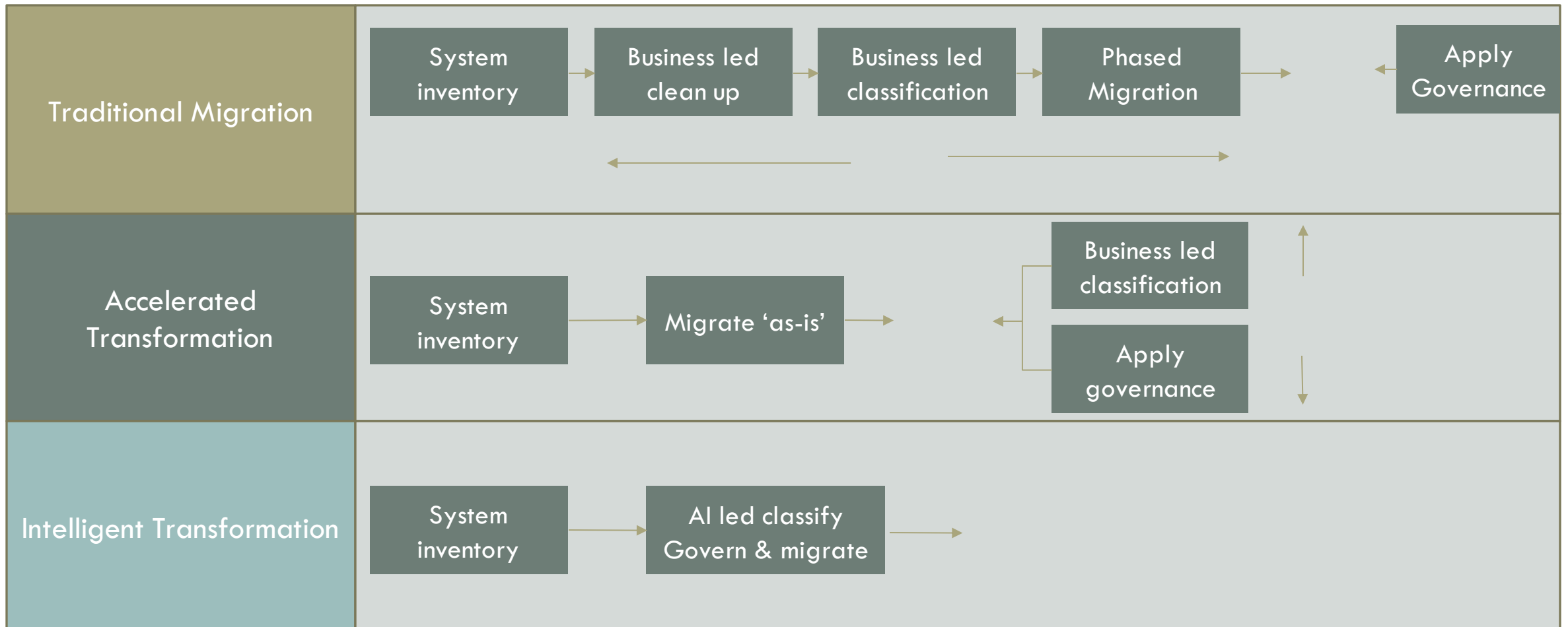


# INFLECTION POINT — M365 EMBRACING TRADITIONAL CONTENT MANAGEMENT ROLES.

From which repositories are you migrating content and documents INTO Office 365/SharePoint Online?	2019	2013
Legacy content management system(s)	33%	29%
Legacy records management system(s)	39%	25%
Line of business applications	35%	31%
SaaS applications	31%	25%
File shares	55%	51%
File sync and share solution	42%	31%
On-premises SharePoint repositories	43%	37%



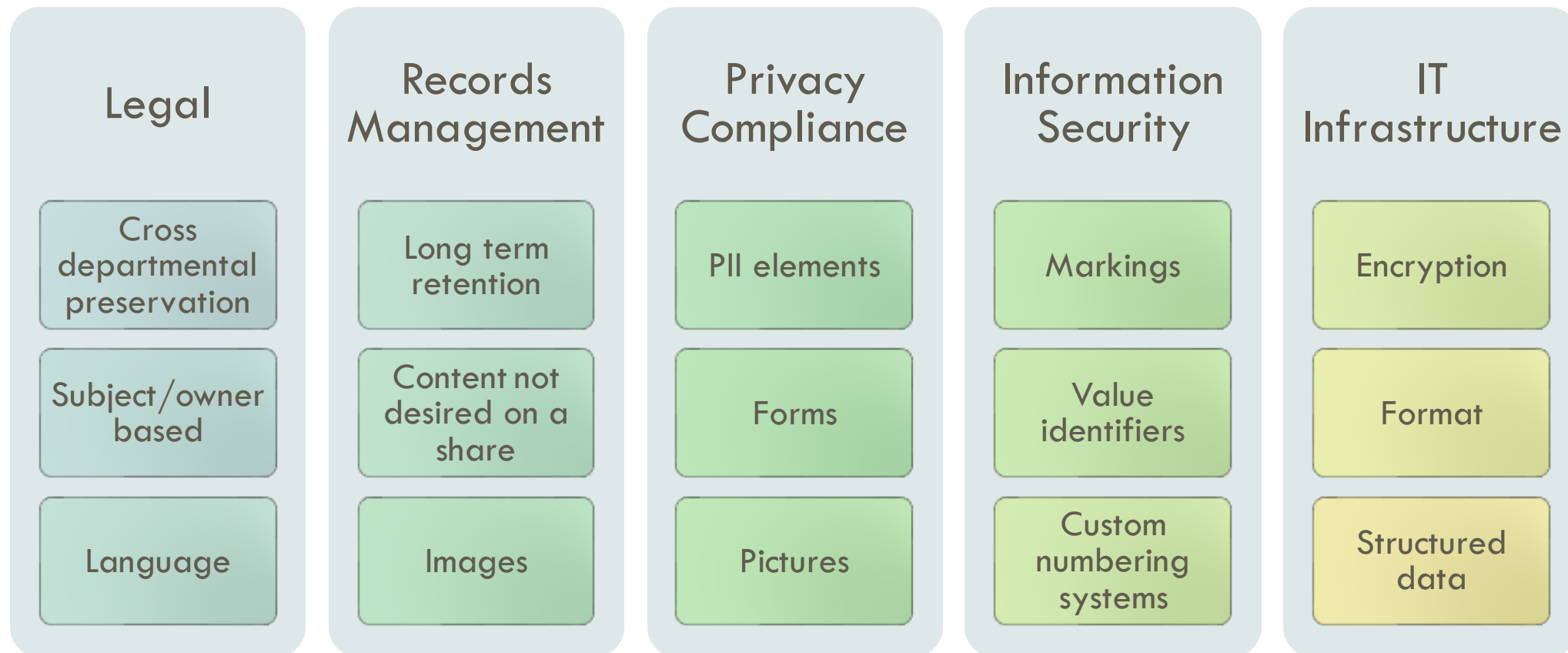
# TRANSFORMATION APPROACHES



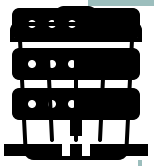
# HOW DO THEY COMPARE?

Approaches	Business Dependency	Migration Speed	Governance Accuracy	Overall Costs
Traditional	High	Subject to business availability (12 months)	High	High
Accelerated	High	High (6-8 months)	Low	Low
Intelligent	Low	High (2-3 months)	High	Medium

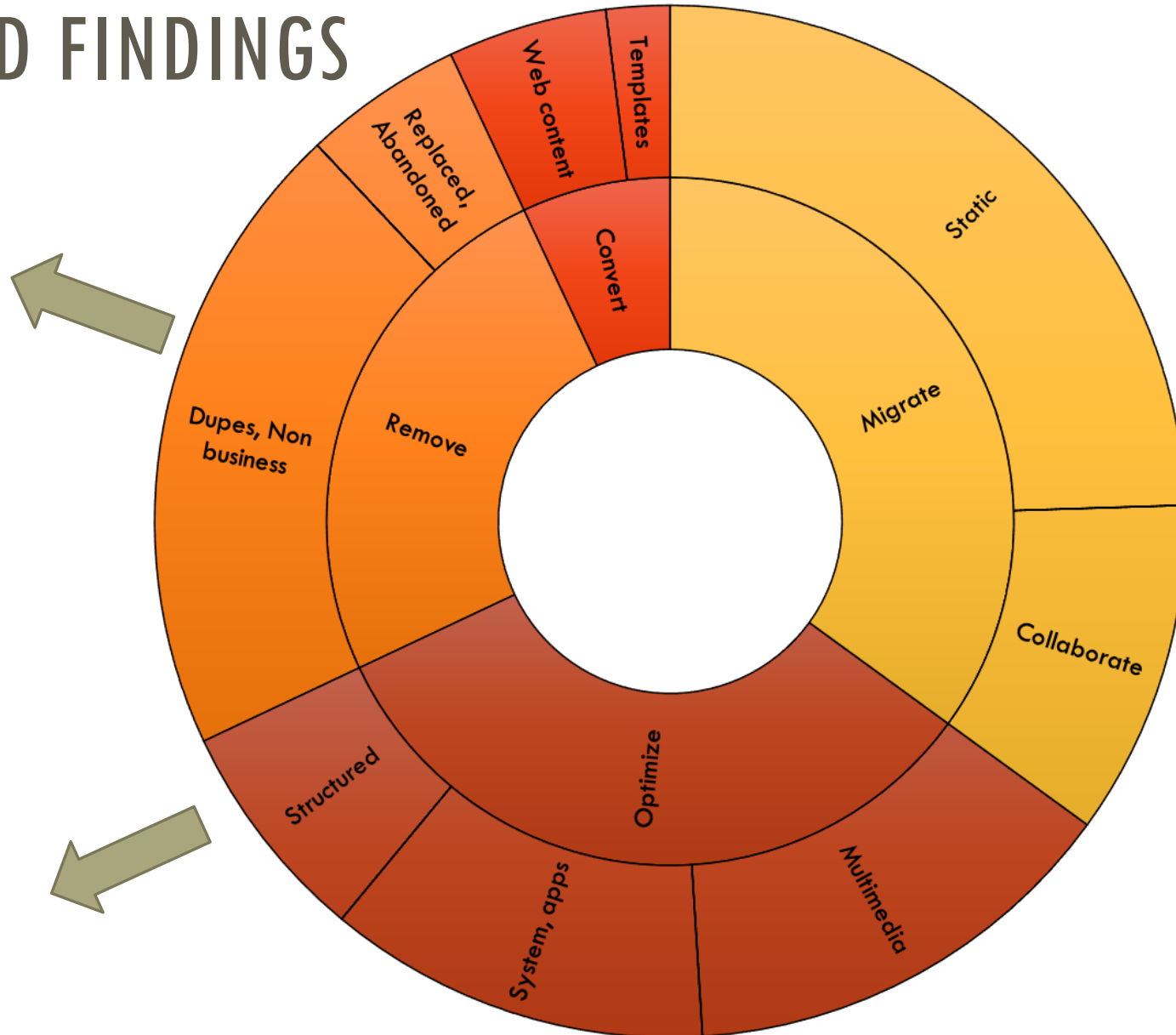
# GOVERNANCE PARTICIPATION/CLASSIFICATIONS



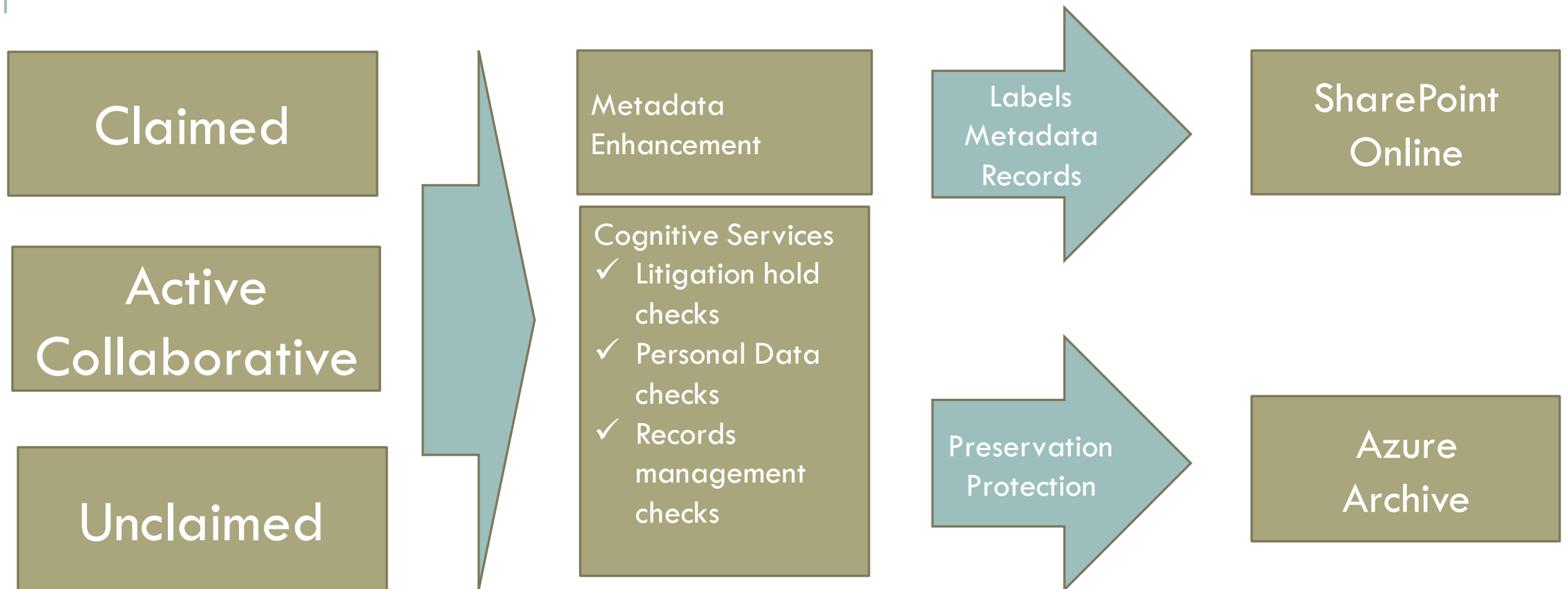




# EXPECTED FINDINGS



# HOLISTIC APPROACH



# KEY PRINCIPLES

<b>Principle 1</b>	The <b>business</b> will have the opportunity to identify important information for capture <b>first</b> and the remaining decisions RIM, Risk, and Legal functions will be informative and additive to those decisions.
<b>Principle 2</b>	Decisions to <b>migrate or eliminate</b> content should be based on <b>validated and authorized rules and definitions</b> rather than individual decisions for consistency and defensibility purposes.
<b>Principle 3</b>	Decision to eliminate, classify, or move content should be made at a single <b>high level of the business</b> rather than the individual level to minimize administrative and duplication of human effort
<b>Principle 4</b>	<b>Content of low value</b> should be <b>remediated before</b> content of high value to minimize the total value of content needing higher scrutiny or classification.
<b>Principle 5</b>	The application of rules will assume (with validation) that <b>attempts in the past to classify content on network shares were intentional</b> and should be considered useful. If a folder is identified, for example, as “Temporary” or “To be deleted” or “draft” or other value indicators it will be treated accordingly.
<b>Principle 6</b>	Wherever possible, the process, and application of rules will assume that good <b>consistency</b> is preferred <b>over</b> complete <b>accuracy</b> .

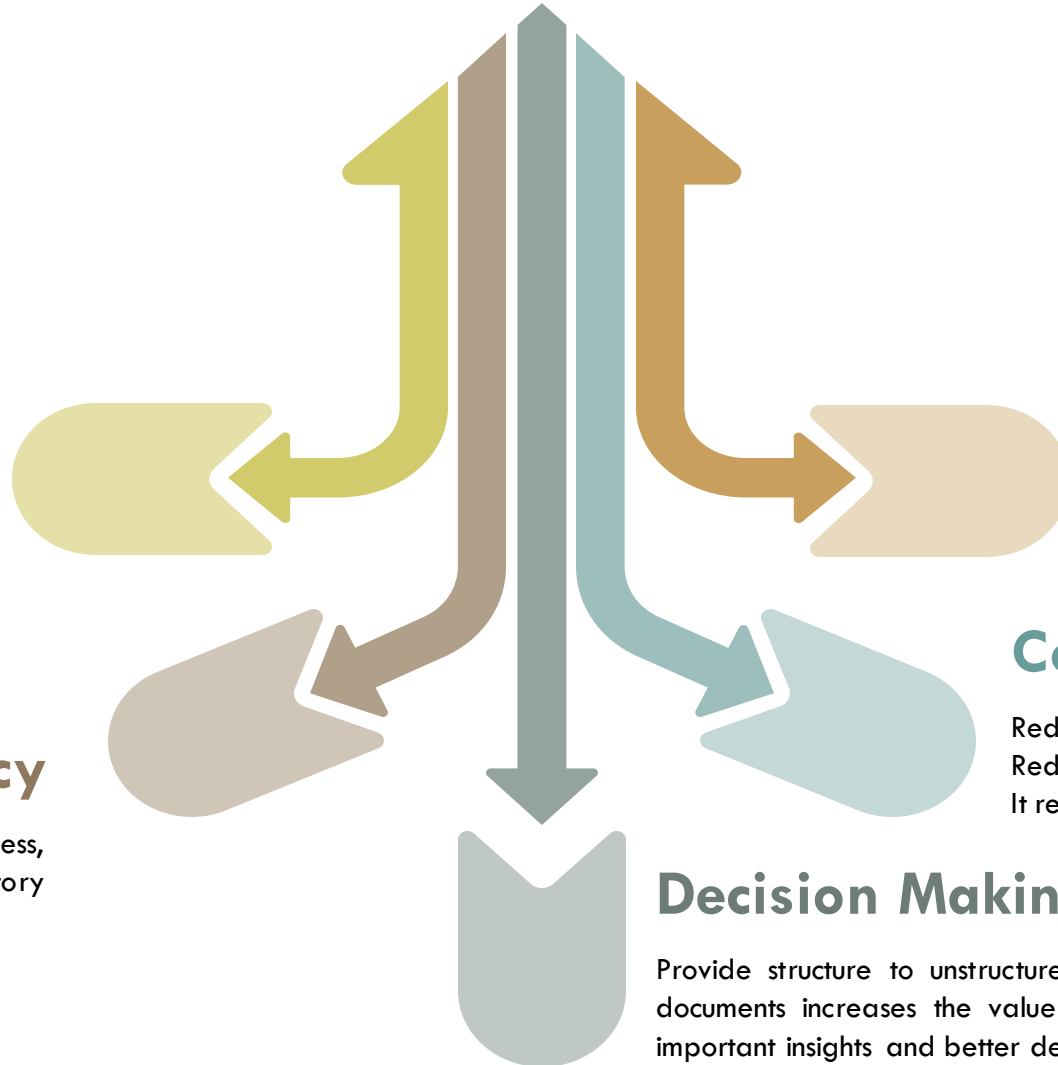
# WHY AUTO-CLASSIFICATION?

## Adapt and Change

Adapt and change to ever increasing and highly variable flow of content. Make content findable rather than just classified

## Transparency

Improvement of compliance, security, responsiveness, user productivity and satisfaction. Increased regulatory transparency and accountability.



## Faster ROI

Quick setup and configuration. Faster, Scalable and objective allowing businesses to efficiently organize content, making it available real time.

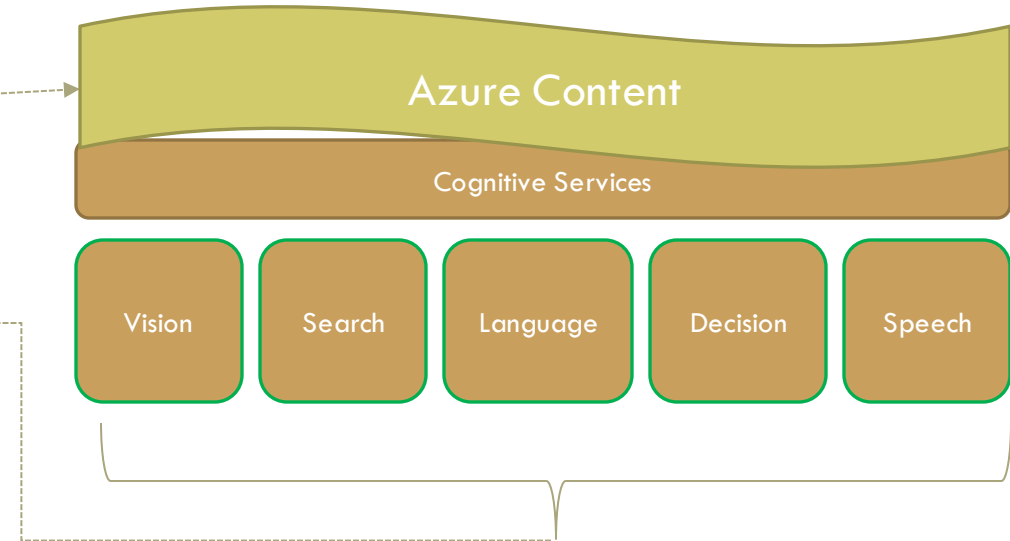
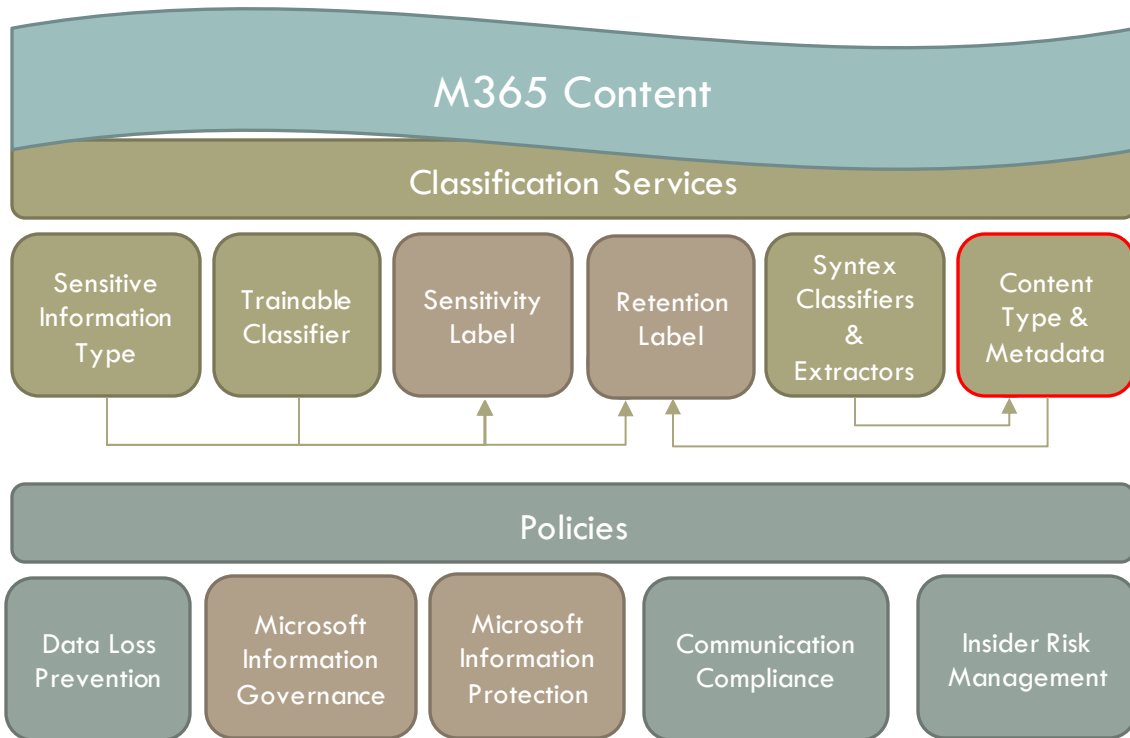
## Cost Savings

Reduced manual intervention, processing and reviews. Reduction of litigation risks/storage costs/eDiscovery costs. It results in a lower overall total cost of ownership

## Decision Making

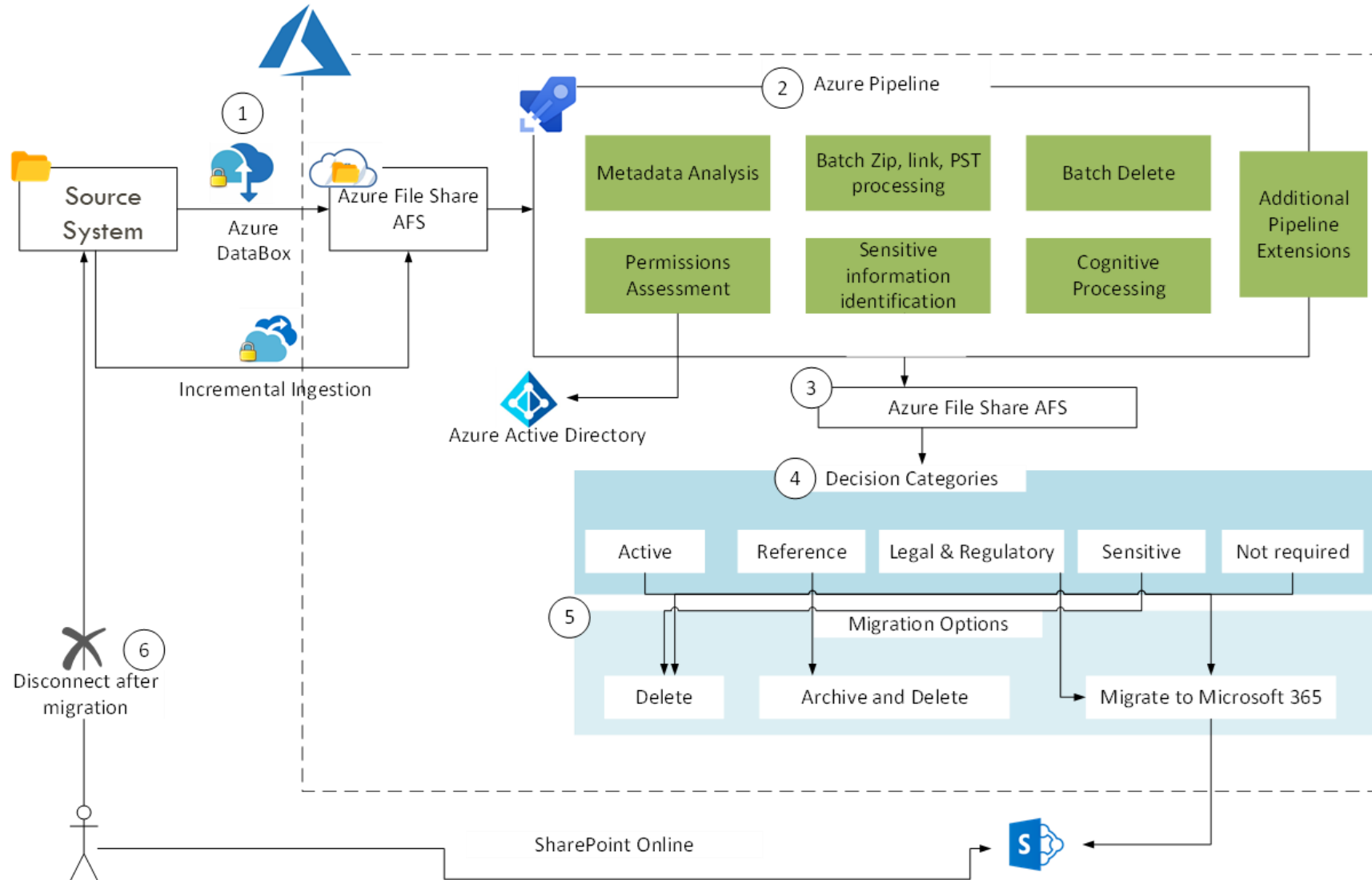
Provide structure to unstructured data. Data from the processed documents increases the value of information helping in gaining important insights and better decision making

# CLASSIFICATION ECOSYSTEM



- OCR
- Regular Expression
- Image recognition
- Translation
- Keyword
- Classifiers
- Forms recognition
- Transcription
- Attribute reference
- Sensitive Information types
- Keyword Query Language

# PRODUCT — INFOTECHTION AUTO-CLASSIFICATION



# INFOTECHTION AUTO- CLASSIFICATION - INTELLIGENT ANALYTICS EXAMPLE

**Share Hierarchy**

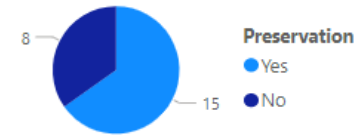
- fileprocessorblob
  - CCN%20sample%20data%20copy%20p...
  - CriticalFileTest.xlsx
  - Different%20Datasets.docx
  - EmailAttachment.docx
  - Excelcontainingdifferentsets.xlsx
  - ExcelCustomSkills.xlsx
  - RE%20Testing%20the%20data.msg
  - Record
    - CriticalFileTest.xlsx
    - Employee%20Contract.docx
    - FOIA\_PIIUsersGuide\_1504.pdf
    - Level2
      - Cambridge%20Case%20Study.docx
      - ExcelCustomSkills.xlsx
      - testing.pst
    - Sample%20data.zip
    - Testing%20the%20data.msg
    - TestingRiskCategory.docx
    - TestingRiskCategory1.docx
    - TestingRiskCategory2.docx
    - TestingRiskCategory3.docx
    - TestingRiskCategory4.docx
    - TestingRiskCategory5.docx
    - TestingRiskCategory6.docx
    - USDA\_PII\_examples.pdf

Files Count  
**23**

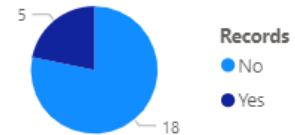
Volume  
**4.11 in MB**



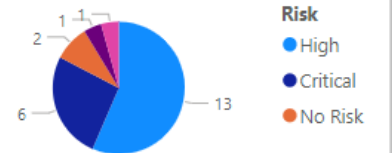
Files by Preservation



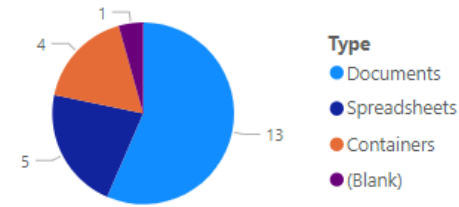
Files by Records



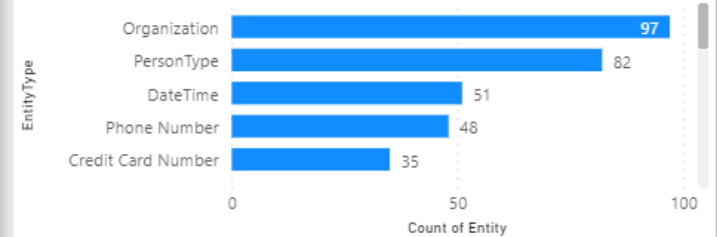
Files by Risk



Files by Type



Entities in Files



FileName	CreationDate	FileLastModified	FileSizeinMB	FileRisk	FileType	ShareFilePath
CriticalFileTest.xlsx	16-06-2021	16-06-2021	0.01	Critical	Spreadsheets	fileprocessorblob/Record/Cri
Different Datasets.docx	25-05-2021	16-06-2021	0.19	Critical	Documents	fileprocessorblob/Different%
EmailAttachment.docx	16-06-2021	16-06-2021	0.01	High	Documents	fileprocessorblob/EmailAttac
Employee Contract.docx	04-06-2021	07-06-2021	0.02	High	Documents	fileprocessorblob/Record/Err
Excelcontainingdifferentsets.xlsx	05-06-2015	15-06-2021	0.01	High	Spreadsheets	fileprocessorblob/Excelconta
ExcelCustomSkills.xlsx	05-06-2015	15-06-2021	0.01	High	Spreadsheets	fileprocessorblob/ExcelCusto
ExcelCustomSkills.xlsx	05-06-2015	15-06-2021	0.01	High	Spreadsheets	fileprocessorblob/Record/le

# INFOTECHTION AUTO- CLASSIFICATION RISK ASSESSMENT EXAMPLE

Showing details for the node: fileprocessorblob

Files with Personal Data **21** | Volume **4.11** in MB

**Entities by Risk**

Entity	Risk Category
No Risk	188
Low	81
Medium	80
High	43

Content Snippet

Do not fax PII to unattended fax machines.

☐ Ensure all printed documents with PII are properly marked with "FOUO -- Privacy Sensitive."

☐ When deemed appropriate, double wrap PII when sending it through the mail. Seal the material in an envelope, address the envelope to an author

ized recipient mark that envelope "FOUO -- Privacy Sensitive"

FileName	Entity	EntityType	RiskCategory	ShareFilePath
Cambridge Case Study.docx	3256-9519-6624	Phone Number	Medium	fileprocessor
Cambridge Case Study.docx	4532-9929-3036-9308	Credit Card Number	High	fileprocessor
Cambridge Case Study.docx	5325-3256-9519-6624	Credit Card Number	High	fileprocessor
Cambridge Case Study.docx	5413-4428-0145-0036	Credit Card Number	High	fileprocessor
Cambridge Case Study.docx	5495-8602-4508-6804	Credit Card Number	High	fileprocessor
Cambridge Case Study.docx	Agnes Nelson	Person	Low	fileprocessor
Cambridge Case Study.docx	Cambridge Analytica	Organization	No Risk	fileprocessor
Cambridge Case Study.docx	early 2018	DateTime	Low	fileprocessor
Cambridge Case Study.docx	Facebook	Organization	No Risk	fileprocessor
Cambridge Case Study.docx	Facebook Inc.	Organization	No Risk	fileprocessor
Cambridge Case Study.docx	FB	Organization	No Risk	fileprocessor
Cambridge Case Study.docx	Julie Renfro	Person	Low	fileprocessor
Cambridge Case Study.docx	Lynette Oyola	Person	Low	fileprocessor
Cambridge Case Study.docx	researcher	PersonType	No Risk	fileprocessor
Cambridge Case Study.docx	Stacey Peacock	Person	Low	fileprocessor
Cambridge Case Study.docx	the day	DateTime	Low	fileprocessor
Cambridge Case Study.docx	University of Cambridge	Organization	No Risk	fileprocessor
Cambridge Case Study.docx	users	PersonType	No Risk	fileprocessor
CriticalFileTest.xlsx	(703) 767-5045	Phone Number	Medium	fileprocessor
CriticalFileTest.xlsx	(703) 767-5045	Phone Number	Medium	fileprocessor
CriticalFileTest.xlsx	15, Hershel Street, New York	Address	Medium	fileprocessor
CriticalFileTest.xlsx	15, Hershel Street, New York	Address	Medium	fileprocessor
CriticalFileTest.xlsx	John Mac	Person	Low	fileprocessor
CriticalFileTest.xlsx	John Mac	Person	Low	fileprocessor
Different Datasets.docx	(555) 555-5555	Phone Number	Medium	fileprocessor
Different Datasets.docx	0 minutes, 21 seconds	DateTime	Low	fileprocessor



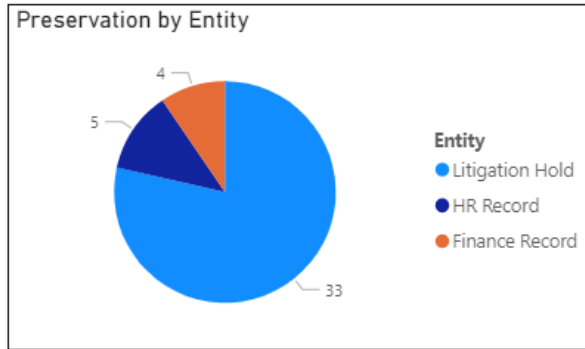
# INFOTECHTION AUTO- CLASSIFICATION LITIGATION ANALYSIS EXAMPLE

Showing details for the node: fileprocessorblob



Files with Preservation  
**15**

Volume  
**1.47** in MB



ContentSnippet

- Sample data/CCWordList.rtf
- adaptation additionality
- albedo
- anoxic event
- Antarctic bottom water (ABW)
- Antarctic oscillation (AAO)
- Antarctica cooling controversy
- anthropogenic
- anthropogenic climate change
- anthropogenic global warming (AGW)
- anti-greenhouse effect
- Arctic dipole anomaly
- Arctic oscillation (AO)
- Arctic shrinkage
- Arctic

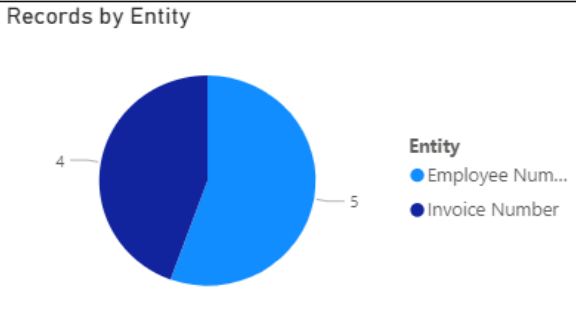
FileName	Entity	EntityType	EntitySubType	RiskCategory	IsRecord
Different Datasets.docx	9735474	HR Record	Employee Number	Medium	Yes
Different Datasets.docx	Climate Change	Litigation Hold	Climate	High	No
Different Datasets.docx	global warming	Litigation Hold	Global	High	No
EmailAttachment.docx	Climate Change	Litigation Hold	Climate	High	No
EmailAttachment.docx	global warming	Litigation Hold	Global	High	No
Employee Contract.docx	12150896	Finance Record	Invoice Number	High	Yes
Employee Contract.docx	4657376	HR Record	Employee Number	Medium	Yes
Excelcontainingdifferentsets.xlsx	46574948	Finance Record	Invoice Number	High	Yes
Excelcontainingdifferentsets.xlsx	9836453	HR Record	Employee Number	Medium	Yes
Excelcontainingdifferentsets.xlsx	Climate Change	Litigation Hold	Climate	High	No
ExcelCustomSkills.xlsx	46584948	Finance Record	Invoice Number	High	Yes
ExcelCustomSkills.xlsx	9836653	HR Record	Employee Number	Medium	Yes
ExcelCustomSkills.xlsx	Climate Change	Litigation Hold	Climate	High	No
ExcelCustomSkills.xlsx	global warming	Litigation Hold	Global	High	No
ExcelCustomSkills.xlsx	46584948	Finance Record	Invoice Number	High	Yes
ExcelCustomSkills.xlsx	9836653	HR Record	Employee Number	Medium	Yes
ExcelCustomSkills.xlsx	Climate Change	Litigation Hold	Climate	High	No
ExcelCustomSkills.xlsx	global warming	Litigation Hold	Global	High	No
Sample data.zip	Climate Change	Litigation Hold	Climate	High	No
Sample data.zip	global warming	Litigation Hold	Global	High	No
Testing the data.msg	Climate Change	Litigation Hold	Climate	High	No
Testing the data.msg	global warming	Litigation Hold	Global	High	No
TestingRiskCategory.docx	Climate Change	Litigation Hold	Climate	High	No
TestingRiskCategory.docx	global warming	Litigation Hold	Global	High	No
TestingRiskCategory1.docx	Climate Change	Litigation Hold	Climate	High	No
TestingRiskCategory1.docx	global warming	Litigation Hold	Global	High	No
TestingRiskCategory2.docx	Climate Change	Litigation Hold	Climate	High	No
TestingRiskCategory2.docx	global warming	Litigation Hold	Global	High	No
TestingRiskCategory3.docx	Climate Change	Litigation Hold	Climate	High	No

# INFOTECHTION AUTO- CLASSIFICATION RECORDS CLASSIFICATION EXAMPLE

Showing details for the node: fileprocessorblob

Files with Records: **5** | Volume: **0.23** in MB

Records by Entity



Entity	Count
Employee Number	5
Invoice Number	4

ContentSnippet

Sheet1  
Employee number 9836653  
Invoice number 46584948  
SSN 568-52-8323  
The paris climate change agreement is signed  
Global warming is on the rise

Sheet1  
Hi, how are you doing today  
Credit Card Number 5495-8802-4508-6804  
Credit Card Number 5495-8602-4508-6804  
Credit Card No 4776-3234-4384-3840  
Employee number 9836453  
Invoice number 46574948  
SSN 568-52-8323  
The paris climate change agreement is signed

FileName	Entity	EntityType	EntitySubType	RiskCategory	IsRecord	ShareFilePath
Different Datasets.docx	9735474	HR Record	Employee Number	Medium	Yes	fileprocessorblob/Different%20C
Excelcontainingdifferent sets.xlsx	46574948	Finance Record	Invoice Number	High	Yes	fileprocessorblob/Excelcontainin
Excelcontainingdifferent sets.xlsx	9836453	HR Record	Employee Number	Medium	Yes	fileprocessorblob/Excelcontainin
ExcelCustomSkills.xlsx	46584948	Finance Record	Invoice Number	High	Yes	fileprocessorblob/ExcelCustomSI
ExcelCustomSkills.xlsx	9836653	HR Record	Employee Number	Medium	Yes	fileprocessorblob/ExcelCustomSI
Employee Contract.docx	12150896	Finance Record	Invoice Number	High	Yes	fileprocessorblob/Record/Emplo
Employee Contract.docx	4657376	HR Record	Employee Number	Medium	Yes	fileprocessorblob/Record/Emplo
ExcelCustomSkills.xlsx	46584948	Finance Record	Invoice Number	High	Yes	fileprocessorblob/Record/Level2
ExcelCustomSkills.xlsx	9836653	HR Record	Employee Number	Medium	Yes	fileprocessorblob/Record/Level2

# INFOTECHTION AUTO- CLASSIFICATION DATA SAMPLING AND MODEL TRAINING EXAMPLE

Showing details for the node: fileprocessorblob

Files

- Select All
- Cambridge Case Study.docx
- CCN sample data copy pwd.xlsx
- CriticalFileTest.xlsx
- Different Datasets.docx
- EmailAttachment.docx
- Employee Contract.docx
- Excelcontainingdifferentsets.xlsx
- ExcelCustomSkills.xlsx
- FOIA\_PIIUsersGuide\_1504.pdf
- RE Testing the data.msg
- Sample data.zip
- Testing the data.msg
- testing.pst
- TestingRiskCategory.docx
- TestingRiskCategory1.docx
- TestingRiskCategory2.docx
- TestingRiskCategory3.docx

Files by Risk

Risk	Count
High	13
Critical	6
No Risk	2

Files by Status

Status	Count
Accept	10
Reject	7
No Risk	6

Random Sampling Review

Accept All  Reject All

File Name	Accept	Reject	Comments
<b>CriticalFileTest.xlsx</b> Risk Category: Critical Preservation: No   Record: No Current Status: Reject Last Status Update By: Rakesh Jain	<input type="radio"/>	<input checked="" type="radio"/>	
<b>Different Datasets.docx</b> Risk Category: Critical Preservation: Yes   Record: Yes Current Status: Accept Last Status Update By: Rakesh Jain	<input checked="" type="radio"/>	<input type="radio"/>	
<b>EmailAttachment.docx</b> Risk Category: High	<input checked="" type="radio"/>	<input type="radio"/>	



ContentSnippet

Do not fax PII to unattended fax machines.










Ensure all printed documents with PII are properly marked with "FOUO -- Privacy Sensitive."


When deemed appropriate, double wrap PII when sending it through the mail. Seal the material in an envelope, address the envelope to an authorized recipient, mark that envelope "FOUO -- Privacy Sensitive," and then place that sealed envelope in a second sealed, unmarked envelope addressed to the authorized recipient.

# INFOTECHTION AUTO- CLASSIFICATION AI BASED TRANSFORMATION AND MIGRATION EXAMPLE

Showing details for the node: fileprocessorblob  

### File Classification Summary

Risk Level	Preservation	Record	File Count	Action
Critical Risk	No	No	2	Delete 
Critical Risk	Yes	No	3	Delete 
Critical Risk	Yes	Yes	1	Delete 
High Risk	No	No	2	Delete 
High Risk	Yes	No	7	Delete 
High Risk	Yes	Yes	4	Delete 
Low Risk	No	No	1	Move to SharePoint 
Medium Risk	No	No	1	Move to SharePoint 
No Risk	No	No	2	Move to SharePoint 

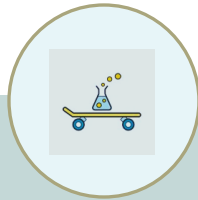


# WHY INFOTECHTION?



## Strong collaboration and co-creation

- A team of Information Governance industry experts with deep knowledge of both technology and the regulations.
- Pathfinders for Governance in Microsoft cloud working with complex multi-geo scenarios
- **Accelerate** the inception and design activities providing time for the business to 'get ready'.



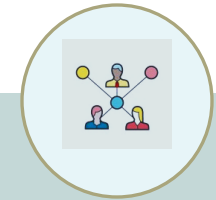
## Agility in the face of technology constraints

- Creating a robust approach for integrating Governance with M365 architecture, **not underestimating** the impact of product configuration on timelines.
- Providing **transparency** and early awareness of issues and constraints to work on mitigating actions.



## Simplified Approach

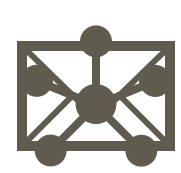
- Weeding out the excess to create a **simple and succinct journey** for user across LOBs.
- Developing an approach **together** to create a journey that reflects strategy.
- Quickly guide – what sticks and what doesn't – and ensure speed of turn-around.



## Proactive management of resistance

- Building **stronger relationships** enabling and empowering business leadership and advocacy of the change.
- Keeping a **close relationship** with Microsoft experts to ensure the simplest route is followed.
- Set up a reference and foundation for data migration and other information related initiatives

Information Governance Practitioners with Hands-on experience of leading from Strategy to Adoption



# THANK YOU

---

VIVEK.B@INFOTECHTION.COM

[HTTPS://INFOTECHTION.COM/](https://infotechtion.com/)