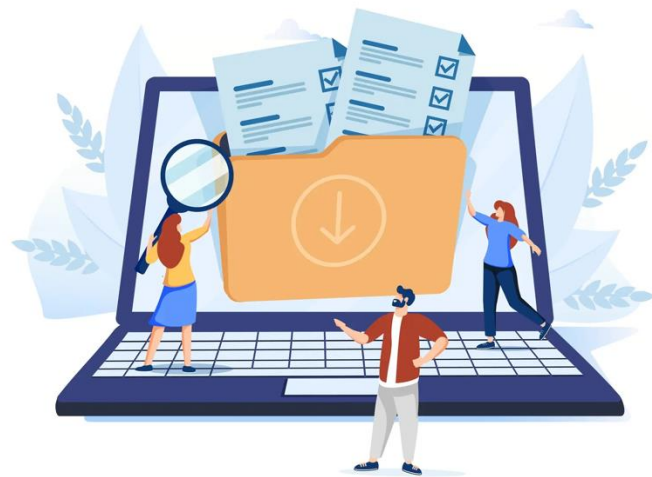




— WEBINAR SERIES —

# Getting Ready for AutoClassification





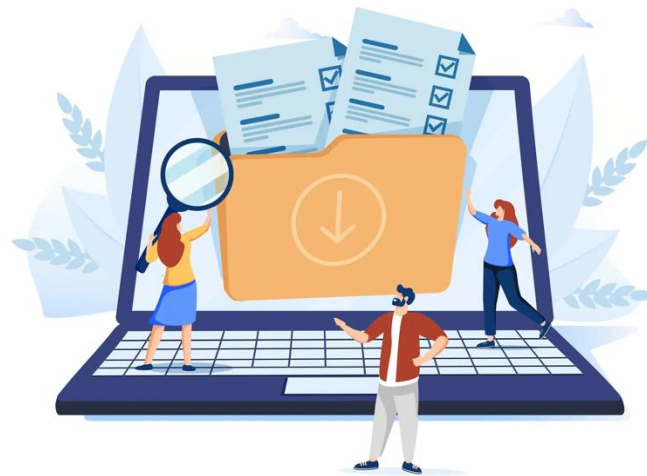
— WEBINAR SERIES —



Sandra Serkes  
Co-Founder & CEO



Jennifer Nelson  
VP Strategic Solutions



# Webinar Series



ROT Remediation:  
file share clean-up  
February 12



Information Governance &  
Enterprise AI Adoption  
June 11



Managing Data Privacy  
while Managing Records  
March 19



AutoClassifying the 3 Rs:  
ROT, Records & Retention  
September 17



The Role of InfoGov  
in AI Readiness  
April 16



Getting Ready for  
AutoClassification  
October 15



AI & AC for Structured  
Data Management  
May 14



Demo Day  
November 12



2026 Trends in IG  
December 10

# About Valora



- Information Governance platform
- Automates discovery, identification, classification & defensible disposition
- Trusted by IG teams, Records & IM teams, Legal, Compliance, IT teams
- Brings multiple enterprise repositories into a single view



# About Valora



## Information Governance / Records Management



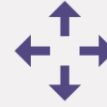
Classification



ROT



Retention



Disposition



Archive

# About Valora



## Enterprise-wide Data Governance



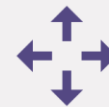
Classification



ROT



Retention



Disposition



Archive



eDiscovery



Legal Hold



Data Privacy



Data Security



Migration



Compliance



Minimization



Provenance



Lineage



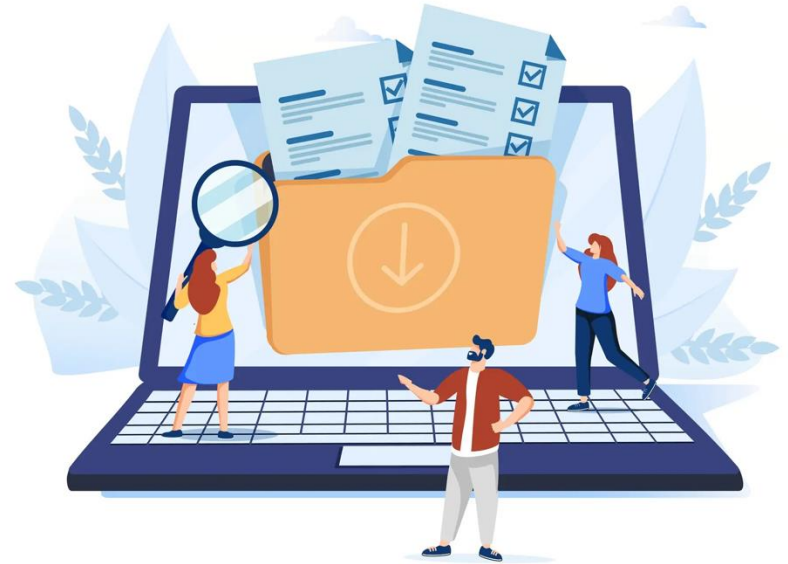
AI Readiness



# Poll Results

---

Who's here today  
& what are your interests?



# Who you are & what your challenges are



Who You Are

Records / Information Management

Knowledge Management

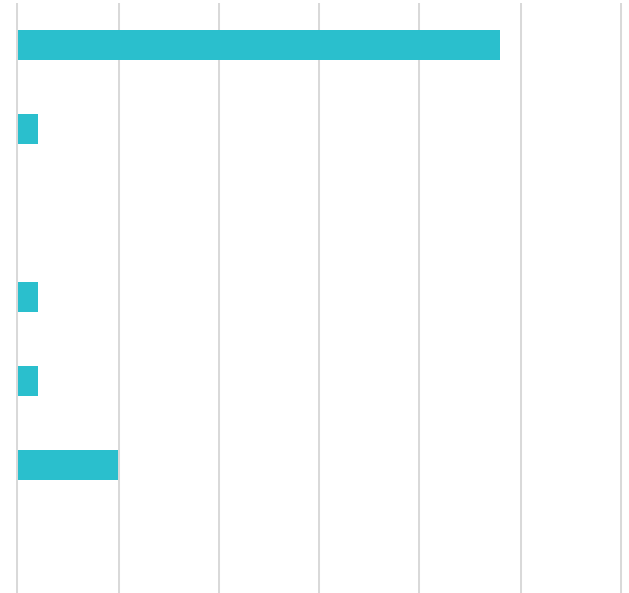
Legal

Compliance

IT

Data Privacy / Security

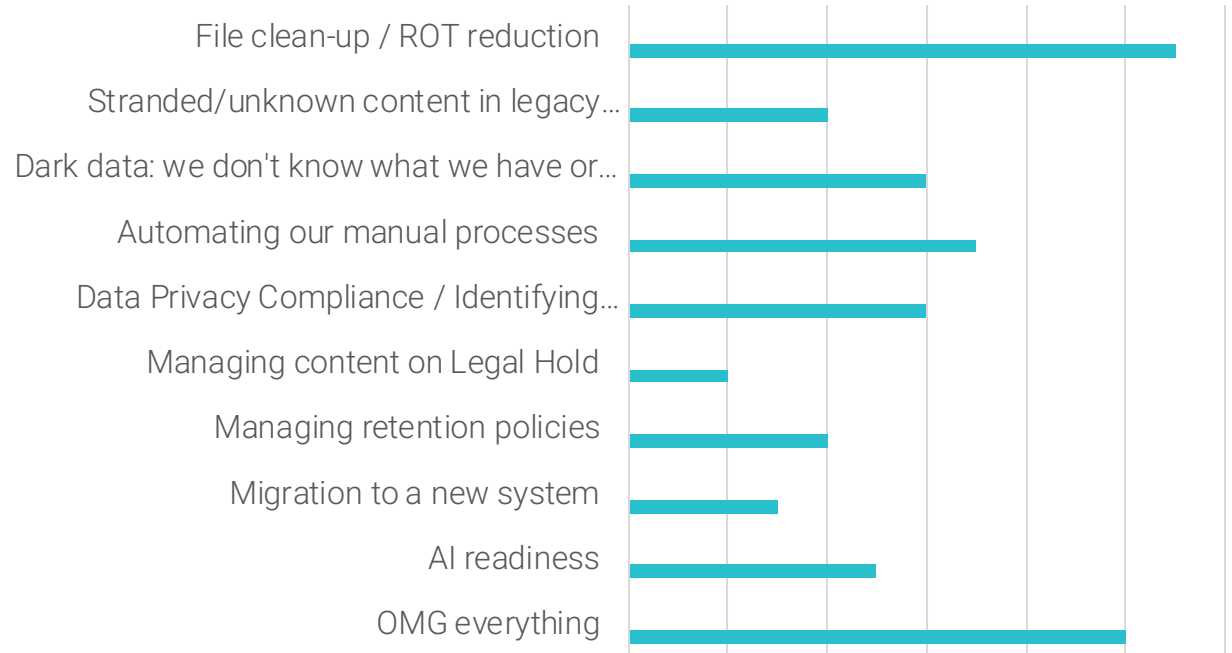
Other



# Who you are & what your challenges are



## Your RIM & IG Challenges



# Housekeeping



Q&A



Recording



Slides



Feedback

# Agenda

AutoClassification 101

Building your Case

Building your Team

Building your Budget

Building your Plan

How to Get Started

# Valora is true AutoClassification



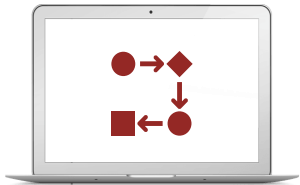
File & Rich Metadata  
Determine Attributes

+



Recognition Algorithms  
Determine DocType

+



Disposition Algorithms  
Rules & Actions

AutoClassification = Rich Metadata + Rules

Software that performs automated analysis & disposition of file/document content



Answers the question:

What is this thing & what do I do with it?

...and then does it.

# File attributes become rich metadata



File & Rich Metadata  
Determine Attributes

+

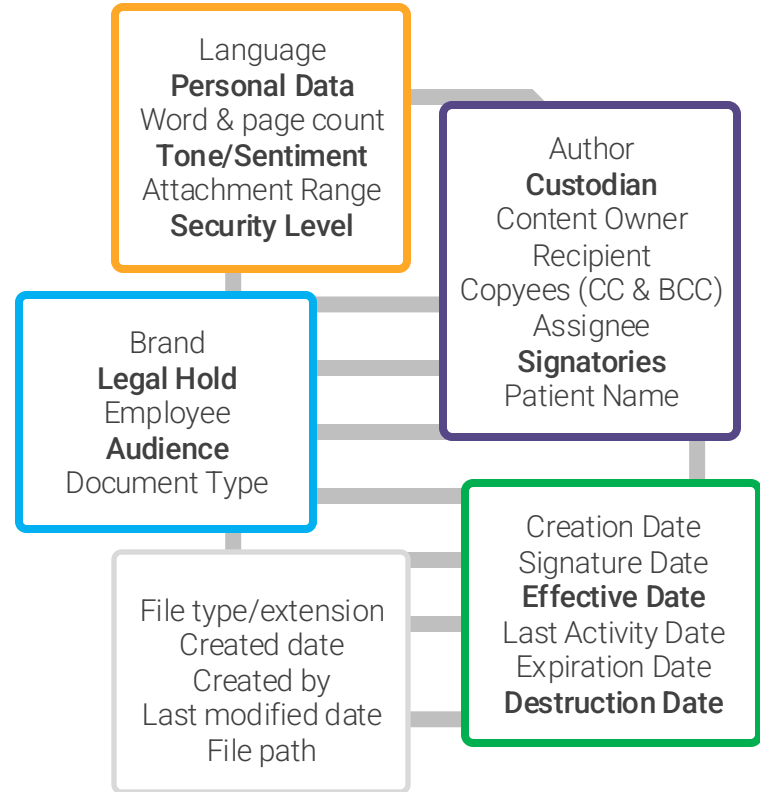


Recognition Algorithms  
Determine DocType

+



Disposition Algorithms  
Rules & Actions



# DocType as defined by RRS



File & Rich Metadata  
Determine Attributes

+

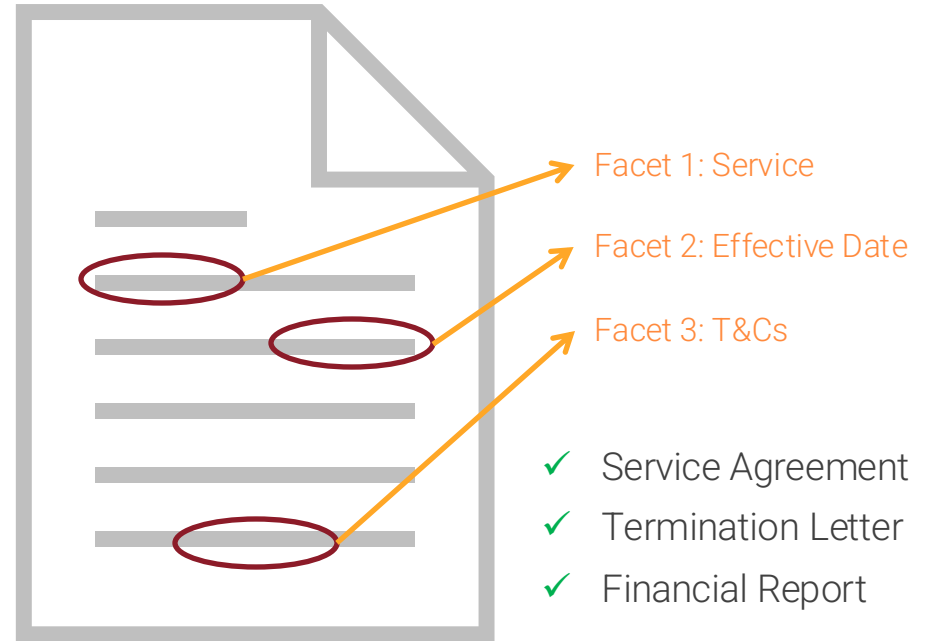


Recognition Algorithms  
Determine DocType

+



Disposition Algorithms  
Rules & Actions



# Defensible Disposition



File & Rich Metadata  
Attributes

+



Recognition Algorithms  
Determine DocType

+



Disposition Algorithms  
Rules & Actions

- Determine the disposition & handling of content
- Typically follow an IF-THEN format, often nested
- Ex: Employment Contract:
  - ✓ Mark as Confidential
  - ✓ Move it to SharePoint
  - ✓ Retain for 8 years after termination
  - ✓ HR approval required before deletion

## Automated, Manual and/or Hybrid Disposition



Tag



Delete



Report



Move



Archive



Hold




Migrate

\*\*\*\*\*

Anonymize

# AutoClassifying Documents

Name	Date modified	Type	Size
 LGPIF Final.pdf	8/13/2016 9:26 AM	Adobe Acrobat Document	171 KB

# AutoClassifying a document



## Local Government Property Insurance Fund Actuarial Services **Contract**

THIS AGREEMENT effective this **1<sup>st</sup> day of February 2005** (the "Effective Date") by and between the Office of the Commissioner of Insurance (the "Agency" or "State of Wisconsin") and **AMI Risk Consultants, Inc.** hereinafter referred to as "Contractor"

The purpose of this Agreement for **actuarial services** is to assist the Agency to accurately assess rate and premium levels and establish loss reserves (incurred but not reported, IBNR) for the Local Government Property Insurance Fund (the "Fund") on an annual basis to ensure the financial stability of the Fund.

**I. Term.** The term of this Agreement shall be for one (1) year, from the 1<sup>st</sup> day of February 2005, and expiring on the 31<sup>st</sup> day of January 2006. The contract may be renewed for two (2) successive one-year periods upon mutual written agreement of the parties. Contractor will notify the Agency six months prior to the annual expiration if it wishes to renew this Agreement and shall specify any amendments the Contractor wishes to propose.

**II. Contact Persons.** For purposes of administering this Agreement, the following representatives of each party are hereby designated as appropriate contact persons:

(a) For the Agency:

Danford Bubolz, Insurance Program Officer  
Local Government Property Insurance Fund  
125 South Webster Street, Madison, Wisconsin 53702

(b) For the Contractor:

Aguedo M. Ingco, President  
AMI Risk Consultants, Inc.  
11410 North Kendall Drive, Suite 208  
Miami, Florida 33176-1031

**III. Actuarial Services Required**

A. Assist the Fund in annually establishing rates and adequate incurred but not reported claims reserves for the Fund.

DocType = Contract  
Effective Date = 2/1/2005  
Party Two = AMI Risk Consultants, Inc.

Term = 1 year  
Renewals = two 1 year terms

Keywords = Actuarial Services

Elsewhere: exhibits,  
amendments, signatories,  
jurisdiction, cover page, etc.



RecordType = Contract  
RecordClass = AP/AR Contract  
Retention = Expiration + 5Y  
Handling = Delete after authorization  
Sensitivity = Confidential  
(protection/redaction based on user access class)



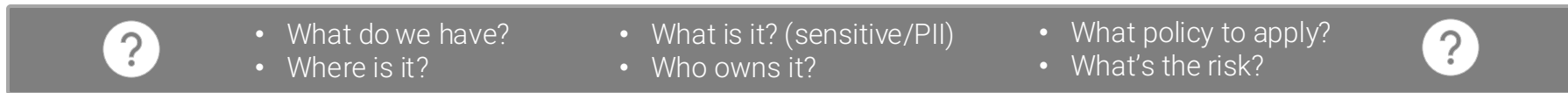
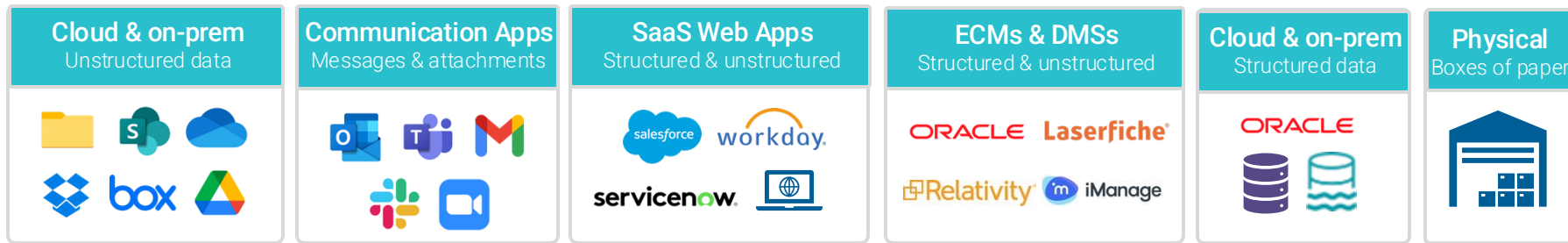
Retain or Delete?

Where? Access?  
How long?


When? How?  
Approval Needed?




# AutoClassifying at the Enterprise Level




**Data Discovery**  
Locate and inventory




**Intelligent Analysis**  
Full-text content analysis



**Auto-Classification**  
Automated classification



**Defensible Disposition**  
Automated rules & workflows



**Delta Monitoring**  
Ongoing scans & audits

- Scans multiple data repositories
- Unstructured/structured
- Cloud & on-prem environments

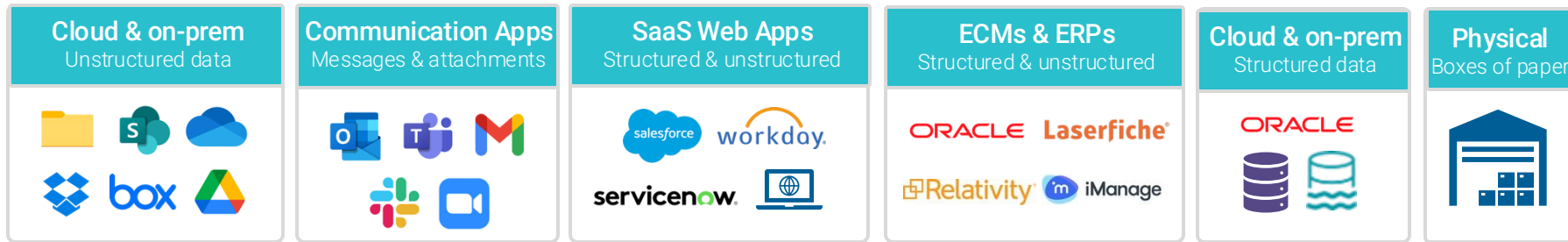
- Full text analysis
- Dark data discovery & inventory
- Text, preview images for analysis

- Classification, based on content
- Rich metadata application
- On-demand/automated reporting

- Automated, Manual, Collaborative
- Automated rules & workflows
- Auditable and traceable

- Scans for new & edited content
- Update changes to policies
- Ensure ongoing compliance

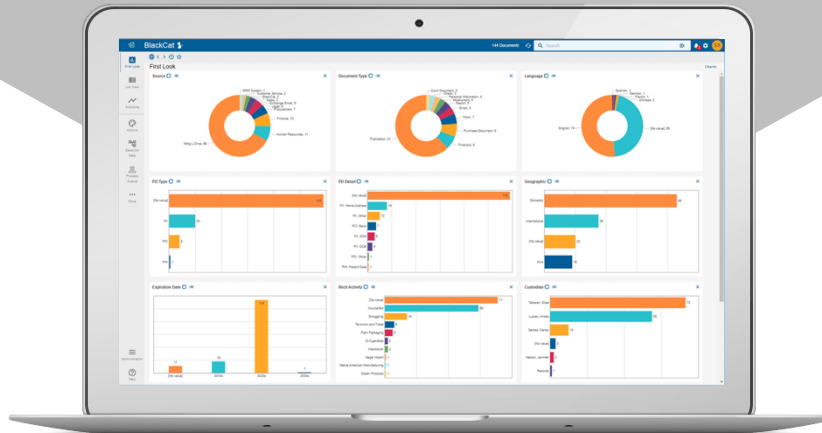
# AutoClassifying at the Enterprise Level



- ✓ What do we have?
- ✓ Where is it?

- ✓ What is it? (sensitive/PII)
- ✓ Who owns it?

- ✓ What policy to apply?
- ✓ What's the risk?



# How Organizations are using AutoClassification



## Data Discovery & Classification

- Identify content: *where* it is and *what* it is to determine *how* and *when* it should be managed.



## ROT & File Clean-up

- Identify, disposition, and defensibly delete Redundant, Obsolete and Trivial data across systems.



## Records Management

- Automate and manage routine and specialized Records handling: M&A, expiration, destruction, migration.



## Data Privacy & GRC

- Locate, identify, redact, manage Personal Data and other corporate sensitive content; DSAR mgmt.



## Legal Hold/eDiscovery

- Reliable data collection for efficient and defensible Legal Hold & eDiscovery processes.



## AI Readiness

- Reduce risk & maximize GenAI initiatives with clean, curated, and trustworthy AI-ready content.

# Benefits of using AutoClassification



## Reduces Errors

Eliminates inconsistencies, human error or oversight



## Reduces Risks

Active responsibility of your data, know what you have, where it is



## Creates Instruction

at the content level, based on metadata & rules



## Defensible

Actions, decisions, movement & access is logged



## Ensures Compliance

Onside with internal policies + regulatory & industry obligations



## Single Pane of Glass

Allows you to see across multiple systems, structured/unstructured



## Easy Data Collection

Any data request: litigation, investigations, DSARs, breaches



## Evergreen

Future-proofs your data. Content is always current, clean



# Building your case

---

Where to start?

# Do I need AutoClassification?



## Multiple Sources

Is my content everywhere?



## Ability to Search

Can I find the content I need?



## Find stuff quickly

Can I access efficiently?



## Unknown data

Is there content I'm missing?



## The right one

Is this the most recent/final?



## Risk

Is this an IG/compliance risk?



## Repeated Search

Performing similar searches?



## Collective Need

Across depts or business units?

# What's the risk?



Internal Policy  
Compliance



Regulatory  
Compliance



Data Privacy  
Compliance

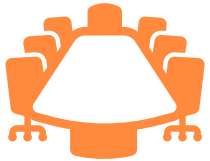


Security  
Compliance



AI / LLM  
Training Models

# What's the risk?



Internal Policy  
Compliance



Regulatory  
Compliance



Data Privacy  
Compliance



Security  
Compliance



AI / LLM  
Training Models



- InfoGov/Records Manager
- Legal/General Counsel
- HR Manager

# What's the risk?



Internal Policy  
Compliance



Regulatory  
Compliance



Data Privacy  
Compliance



Security  
Compliance



AI / LLM  
Training Models



- Compliance Team
- CFO / CISO

# What's the risk?



Internal Policy  
Compliance



Regulatory  
Compliance



Data Privacy  
Compliance



Security  
Compliance



AI / LLM  
Training Models



- DPO / CIO
- GDPR/CCPA

# What's the risk?



Internal Policy  
Compliance



Regulatory  
Compliance



Data Privacy  
Compliance



Security  
Compliance



AI / LLM  
Training Models



- CSO / CISO
- IT

# What's the risk?



Internal Policy  
Compliance



Regulatory  
Compliance



Data Privacy  
Compliance



Security  
Compliance



AI / LLM  
Training Models



- CIO / Transformation
- IT / Data Science

# Common thread?



Internal Policy  
Compliance



Regulatory  
Compliance



Data Privacy  
Compliance



Security  
Compliance



AI / LLM  
Training Models

- ✓ What is this piece of content?
  - ✓ Where should it live?
  - ✓ Who should have access to it?
  - ✓ How long should we keep it?
- ✓ How do we retrieve it should we need it?

# What's the biggest priority?



Internal Policy  
Compliance



Regulatory  
Compliance



Data Privacy  
Compliance



Security  
Compliance



AI / LLM  
Training Models

? Identify the greatest risk and lead with that ?

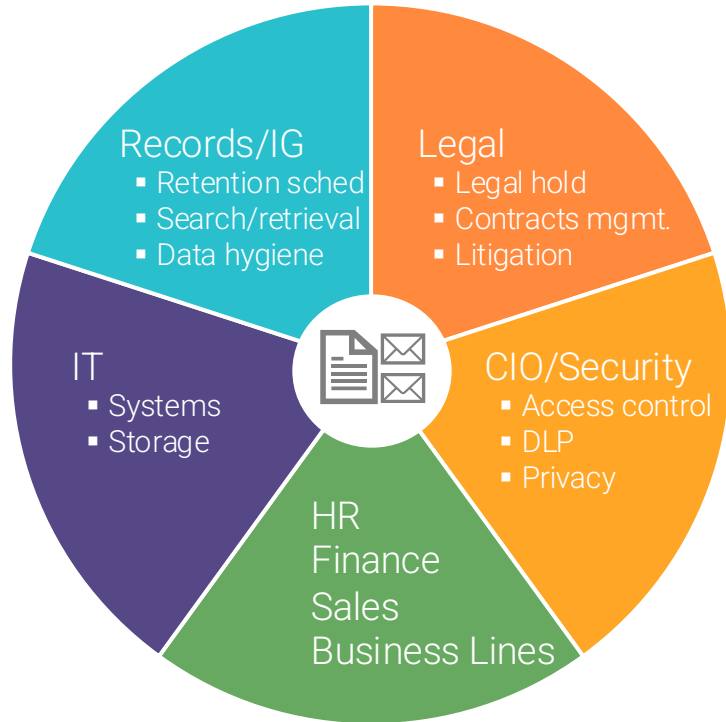


# Building your team

---

Who you need in your court

# Same content, different goals



Every piece of content created has different value to each group

Who owns the content?

Whose goal is most important?

What poses the greatest risk?

Who owns the budget?

Common Goal:

each group needs to be compliant with their mandate

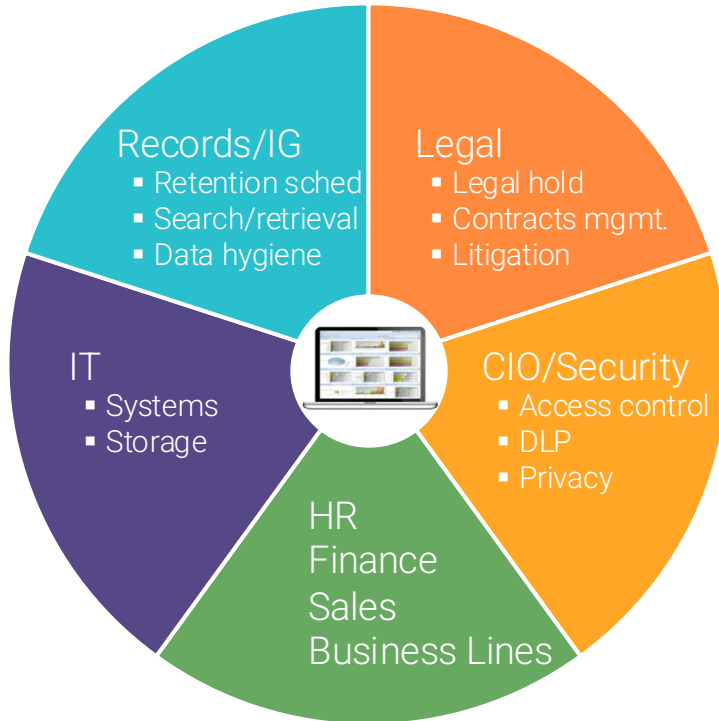
# ARMA Information Governance Implementation Model



## STEERING COMMITTEE

Information Governance Leadership  
Business Units  
Information Technology  
Legal  
Risk/Compliance  
Privacy  
RIM  
Information Security

# Shared content, shared goals



Content management is everyone's issue

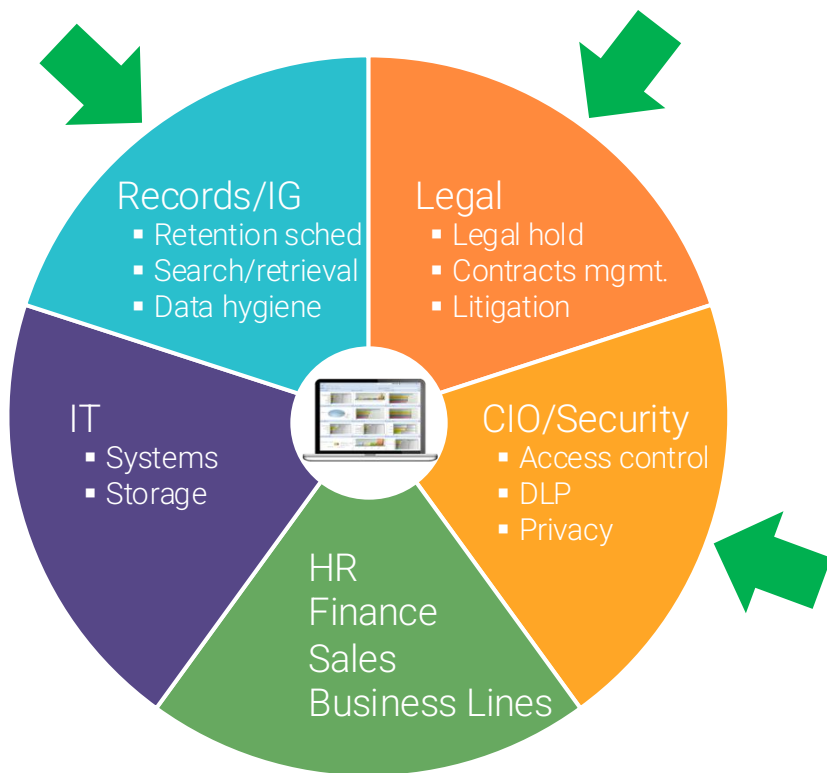
Organization and control are for the greater good

Should be a shared resource

Shared responsibility, shared cost

Common Goal:  
centralized management  
with *distributed benefits*

# Where to start?



Teams you know have the same issues you do

- ✓ What do we have?
- ✓ Where it is?
- ✓ What is this piece of content?
- ✓ Where should it live?
- ✓ Who should have access to it?
- ✓ How long should we keep it?

# Front Line



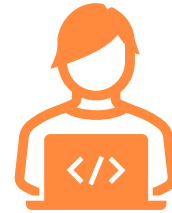
IG

Knows  
how to manage



SME

Knows  
what it is



IT

Controls access



# Building your budget

---

Realizing ROI

# Budget



Tools



People



Time

# ROI Calculator – ROT

*To use this ROI Calculator, enter your actual values into the yellow highlighted cells.  
The rest of the cells will auto-calculate for you.*

		<b>Notes</b>
Estimated Total TB (uncompressed)	250	
Typical active content hosting fees (cloud hosting)	\$0.35	per GB per month
Estimated annual content hosting fees currently (before AutoClassification)	\$1,050,000	
Estimated amount of ROT (Dupes, Past Retention, Junk Files, etc.)	50%	
Estimated TB for hosting After AutoClassification	125	Assumes ROT removed
Estimated annual content hosting fees After AutoClassification	\$525,000	
<b>Total Annual Cost Savings for ROT Removal</b>	<b>\$525,000</b>	
1-year cost of Valora AutoClassification solution	\$276,000	Enterprise level (ROT only)
<b>Total Return on Investment (ROI)</b>	<b>90%</b>	Per Year

# ROI Calculator – Records/Retention Management

		Notes
Estimated Total TB (uncompressed)	250	
Files/TB (average)	3,000,000	
Estimated Total Number of Files	35,000,000	
Total number of files/records tagged and classified for retention per person-hour of effort	120	Assumes 2 files/records per minute (30 seconds each)
Total person-hours to manually classify full corpus	291,667	
Average annual salary of person performing records classification & retention efforts	\$80,000	
Loaded cost of labor (includes overhead, benefits, supervision, etc.)	\$104,000	
Estimated total cost of corpus record classification effort Before AutoClassification	\$ 15,166,667	
Years to AutoClassify corpus manually	7	Enterprise level (Records only)
3-year cost of Valora AutoClassification solution	\$1,044,000	Enterprise level (Records only)
<b>Total Cost Savings for Retention</b>	<b>\$ 14,122,667</b>	
<b>Total Return on Investment (ROI)</b>	<b>1353%</b>	Per Year

# ROI Calculator – eDiscovery

Estimated number of files for Processing, Hosting & Doc Review in a typical matter	350,000	
Estimated % of produced files to opposing (post review)	5%	
Estimated number of produced files to opposing	17,500	
Number of files not produced	332,500	
Cost per document to perform Processing, Doc Review & Hosting	\$1.50	Includes culling, processing, review and hosting
Cost per matter of files not produced	\$498,750	
Number of litigation matters per year	8	
Estimated annual cost of performing eDiscovery tasks on non-produced files	\$3,990,000	
eDiscovery cost reduction % from using AutoClassification ahead of matters	50%	Savings come from pre-tagging files so they are not collected for DocReview and later discarded.
<b>Total Annual Cost Savings for eDiscovery</b>	<b>\$1,995,000</b>	
1-year cost of Valora AutoClassification solution	\$264,000	Starter Level (Total Governance)
<b>Total Return on Investment (ROI)</b>	<b>756%</b>	Per Year

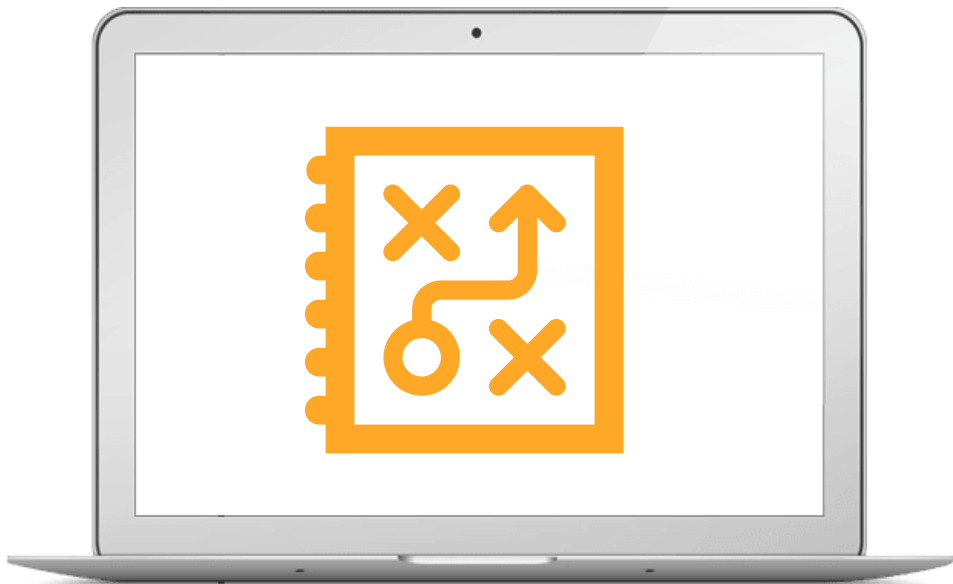
# ROI Calculator – Privacy / Data Requests

Total person-hours to perform one search & retrieval effort for data requests (eDiscovery collection, FOIA or DSAR requests, KM searches, past retention pulls, investigations, etc.)	80	
Total number of search & retrieval efforts performed per month across org	20	
Average annual salary of person performing search & retrieval efforts	\$80,000	
Loaded cost of labor (includes overhead, benefits, supervision, etc.)	\$104,000	
Total cost of each search & retrieval effort	\$4,160	
Estimated total annual cost of performing search & retrieval efforts per year Before AutoClassification	\$998,400	
Estimated total annual cost of performing search & retrieval efforts per year After AutoClassification	\$0	Assumes responses < 1 minute and/or requestors self-serve
<b>Total Annual Cost Savings for Search &amp; Retrieval</b>	<b>\$998,400</b>	
1-year cost of Valora AutoClassification solution	\$348,000	Enterprise level (Privacy only)
<b>Total Return on Investment (ROI)</b>	<b>187%</b>	Per Year

# ROI Calculator – Cost Sharing

*This ROI factors in savings from all four areas sharing the same base processing system across use cases.*

	<b>ANNUAL SAVINGS</b>	<b>Notes</b>
<b>ROT REMOVAL CONTRIBUTION</b>		
Total Annual Cost Savings from ROT removal	\$525,000	
<b>RECORDS RETENTION CONTRIBUTION</b>		
Total Cost Savings for Retention	\$2,017,524	
<b>DATA PRIVACY CONTRIBUTION</b>		
Total Annual Cost Savings for Search & Retrieval	\$998,400	
<b>eDISCOVERY CONTRIBUTION</b>		
Total Annual Cost Savings for eDiscovery	\$1,995,000	
<b>ALL SAVINGS COMBINED</b>		
Savings across all areas impacted	\$5,535,924	
1-year cost of Valora AutoClassification effort	\$396,000	Enterprise level (Total Governance)
<b>Total Return on Investment (ROI)</b>	<b>1298%</b>	Per Year



# Building your plan

---

Implementation & execution

# Eating the Elephant

- ✓ Needs assessment
  - ✓ Identified the risks
  - ✓ Got your team / stakeholders
  - ✓ Budget
- ☐ Set priorities
  - ☐ Content location or custodian
  - ☐ Which group to start with

# Important Topics to Solidify



## Implementation & Migration Plans

- Where is best location for solution – on prem or cloud?
- Going forward, which systems will be used as short-term/current content holders?
- What content will be migrating? When? In what sequence or priority?



## Long-term content storage/management plans

- Which systems will be used as long-term/ records/ preservation holders?
- Can we ensure cloud access to network data stores?
- What connectors will be needed? What APIs available?



## Content Attributes

- What data facets are important to us now and in the future? What results?
- How much total content (prior, now and accrual rates)?
- What content types? What changes/additions over time? What languages present?



## Internal Operations

- What is ideal request/query workflow? What self-serve plans?
- Looking back, what has been the usage/query/need for people to go back to content stored in locations? Who? How often? What results?

# Second Level Considerations



## Overcoming gotchas, anticipating issues/objections

- Initial cost vs. ROI over time
- Time-to-implement
- People, training and evangelization
- Finding other use cases outside of the common
- Trying to use IT tools to solve the challenge



## Where things can go wrong

- Over-promising results without proof-of concept exercises
- Under-estimating the machine learning process
- Obligations for time commitments from tangential players
- Having poor or myopic views on your data



## Looking beyond RIM & eDiscovery and building real business value

- Policies and actionable procedures to implement and sustain
- Meaning-based AC & Rules, AC Renaissance
- Getting past folder taxonomies for managing/retrieving content
- Managing the content estate like financial portfolio

# Project Planning & Readiness



Build your case

- ✓ Needs assessment
- ✓ What's the risk?



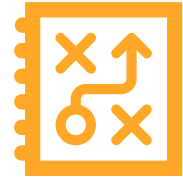
Build your team

- ✓ Stakeholders
- ✓ Collaboration



Build your budget

- ✓ Find savings
- ✓ Path to quick ROI



Build your plan

- ✓ Prioritize objectives
- ✓ Phased roll-out

# How to Get Started (how to get internal buy-in)

## Pilot Use Case

Start with a compelling use case

- Quick and easy that will prove ROI quickly
- 1 content owner/ROT?
- Prove ROI



## Look Forward

Always have an eye on future use cases

- What do you need for this application or use case?
- What will you need for future use cases?



## Measure Results

Benchmark results and accuracy of pilot

- Accuracy of results?
- Increase guidance or seek SME?
- What reporting was useful? What else is needed?



## Build on Success

Build business case on pilot alone then expand

- Prove ROI functioning system
- Where to next?
- Additional use cases running on same platform



## Spread the Word

Sharing is caring, expand internally

- Host internal showcases for mgmt & other departments
- Design pilots that highlight important content, results or exposure points



# Cool. How long is that going to take?



## Complexity

Of your data environment: cloud, on-prem, structured, unstructured data environments



## Goals & Requirements

Input from stakeholders for enterprise-wide approach to data governance.



## Amount of Data

Smaller organizations with less data will be faster to process than large enterprise.

Weeks

Months

# Phased Implementation



## Planning

Prioritize data environments, align stakeholders, set KPIs



## Proof of Concept

Limited data set to benchmark and forecast scope/scale



## Implementation

Implement scale to process all enterprise content



## Delta Monitoring

Continuously look for and process new & edited content

Weeks

Months

# Are you ready for AutoClassification?



- ✓ Identified starter/pilot data set for analysis
- ✓ Core team is ready for questions / decisions / evaluations
- ✓ Guidance data is ready (ex: RRS, taxonomy, sensitivity schema)
- ✓ Project management 101: KPIs outlined, timeline established
- ✓ Executive commitment / Confirmed Budget
- ✓ Vendor(s) are vetted

# The kinds of things we will ask you

- ✓ Infrastructure “bill of lading”
- ✓ Service accounts & access permissions
- ✓ Target data sets for analysis
- ✓ Guidance data location & access

## Recommended Configuration for Valora Technologies PowerHouse and BlackCat

### Amazon Web Services (AWS)

#### Starter

Purpose	Recommendation	Count
<b>Development Environment</b>		
<b>PowerHouse</b> Coordination Server and classification processor	EC2 <b>r6a.large</b> (2 vCPU, 16 GB RAM), <b>100 GB</b> local EBS storage General Purpose SSD (gp3)	1
<b>BlackCat</b> Controller, web server, and Elasticsearch 1	EC2 <b>r6a.xlarge</b> (4 vCPU, 32 GB), <b>150 GB</b> local EBS storage General Purpose SSD (gp3)	1
<b>BlackCat</b> Elasticsearch 2, 3	EC2 <b>r6a.large</b> (2 vCPU, 16 GB RAM), <b>150 GB</b> local EBS storage General Purpose SSD (gp3)	2
<b>Production Environment</b>		
<b>PowerHouse</b> Coordination server	EC2 <b>r6a.large</b> (2 vCPU, 16 GB RAM), <b>100 GB</b> local EBS storage General Purpose SSD (gp3)	1
<b>PowerHouse</b> Classification processors	EC2 <b>r6a.large</b> (2 vCPU, 16 GB RAM), <b>64 GB</b> local EBS storage General Purpose SSD (gp3)	5
<b>BlackCat</b> Controller and processor	EC2 <b>r6a.large</b> (2 vCPU, 16 GB RAM), <b>100 GB</b> local EBS storage General Purpose SSD (gp3)	1
<b>BlackCat</b> Web server	EC2 <b>r6a.large</b> (2 vCPU, 16 GB RAM), <b>64 GB</b> local EBS storage General Purpose SSD (gp3)	1
<b>BlackCat</b> Elasticsearch	EC2 <b>r6a.large</b> (2 vCPU, 16 GB RAM), <b>150 GB</b> local EBS storage General Purpose SSD (gp3)	3
<b>Shared</b>		
Network storage	<b>FSx</b> for Windows File Server 1024 GB w 32 MB/s throughput	1

# The kinds of things we will ask you

- ✓ Scoping & tagging questions
- ✓ Special & edge cases
- ✓ Disposition & actioning
- ✓ Reporting needs
- ✓ Workflow, authorizations & auditing

## ROT Detail

Field Value	Examples	Value Type (Standard/Custom)
Obsolete: File Extension	File has extension from known obsolete list (ex: .tmp)	Standard
Obsolete: Past Retention	File Expiration Date is before the current date	Standard
Obsolete: Temporary or Draft	File says [DRAFT] in title or body	Standard
Obsolete: Version	File is a prior version from others, which are later (in sequence, numbering or date)	Optional
Obsolete: File Naming	File has words like "old", "delete", or "temporary" in filename	Optional
Redundant: Identical Dupe	File is an identical dupe of at least one other file	Standard
Trivial: Auto-Reply	File is an auto-reply (Ex: out of office email)	Standard
Trivial: Low Business Value	Content has no discernible business value, and indications of non-value (ex: "Happy Birthday" in content, DocType=recipe)	Standard
Trivial: Software Code or Executable	File is a piece of some other software program	Standard

# The kinds of things we will ask you

- ✓ Meeting KPIs
- ✓ Work & deliverables acceptance
- ✓ Data quality & results
- ✓ Suitability to purpose
- ✓ Recommendations on moving forward



## User Acceptance Test Form Functionality

Project Name	
Test Date & Time	
Release	
Tester	
Area	Item
Login & Setup	User can successfully login w/ given UN & PW
Login & Setup	User can change PW
Login & Setup	User can logout
Login & Setup	Forgot username/ forgot PW buttons work
Login & Setup	Auto-logout after 60 minutes
Login & Setup	Hover text is present
Help	System Help is displayed under "?" button
Help	Instance Help is displayed under "i" button
Help	Can flip between help pages
Help	Links to particular help pages and videos function
Alert	Create alerts: This file must have the headers: timestamp,type,message,url,info. Valid types are Report, Maintenance, and blank. Timestamp can be either US or ISO and the time is optional. URL is optional.
Alert	Alerts will have the blue dot removed after it is read and the date has today or past
Alert	Able to see all previous alerts
Data Values	Make sure field values are imported correctly (ie ROT, PD fields)
FirstLook	All charts display (may need refresh)
FirstLook	Charts displayed match spec
FirstLook	User can change set of displayed charts



[www.valoratech.com](http://www.valoratech.com)



Book a Demo



Webinar Series



[/valora-technologies](https://www.linkedin.com/company/valora-technologies)

## Q&A

---

Thank you



**Sandy Serkes**  
President & CEO  
[sserkes@valoratech.com](mailto:sserkes@valoratech.com)



**Jennifer Nelson**  
VP Strategic Solutions  
[jnelson@valoratech.com](mailto:jnelson@valoratech.com)