



Managing Records Retention & Data Minimization with AutoClassification

Webinar Mash-up



Sandra Serkes
Co-Founder & CEO



Jennifer Nelson
VP Strategic Solutions

Agenda

Introduction to Valora

What is AutoClassification?

AutoClassification for Records Mgmt

What is Data Minimization?

Where do DM and RIM/IG intersect?

How to Get Started

Valora Technologies

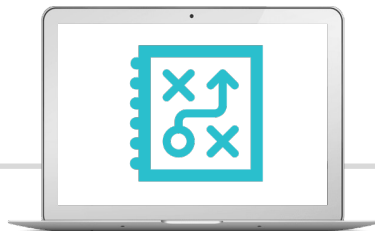
AutoClassification platform used by
Data Governance, Records Managers, Legal, Compliance & IT teams
to locate and automate the identification, classification & defensible disposition
of content across the enterprise data estate.



Technology

Enterprise-grade SaaS platform

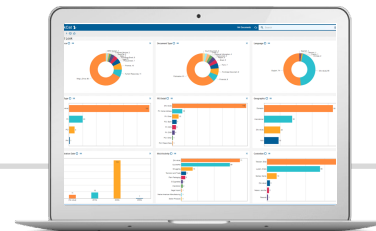
- Flexible, scalable, customizable
- True AutoClassification engine



Methodology

Proven Process & Best Practices

- 20+ year pioneers in InfoGov
- In-house Professional Services team



360° View

Single pane of glass

- Un/structured, on-prem/cloud, SaaS
- ID, classify, defensibly disposition

Agenda

Introduction to Valora

What is AutoClassification?

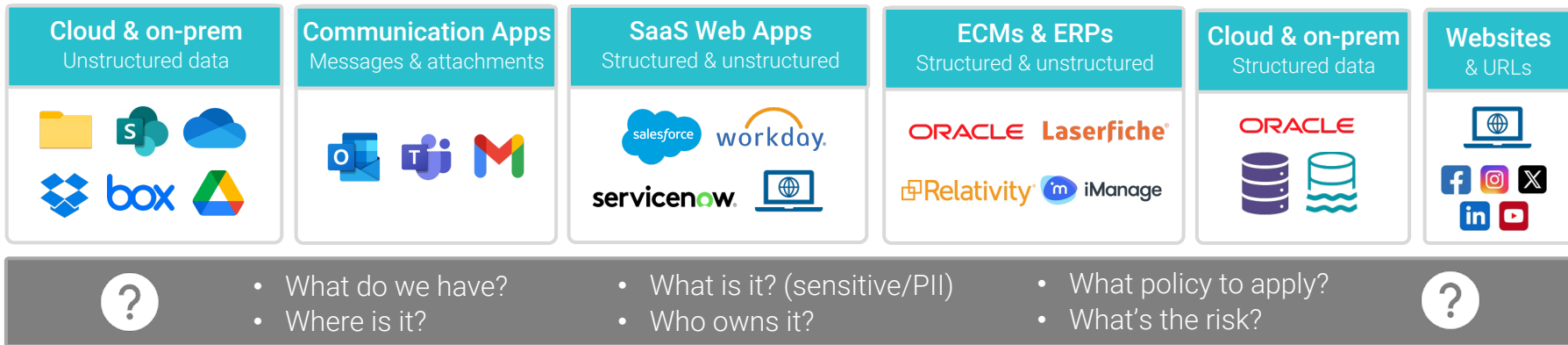
AutoClassification for Records Mgmt

What is Data Minimization?

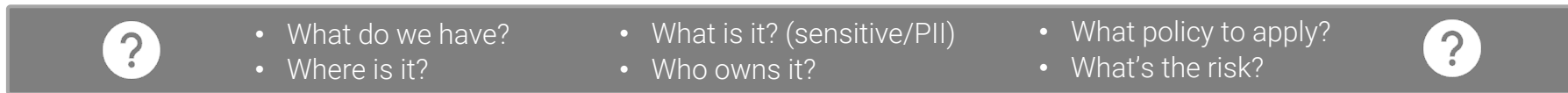
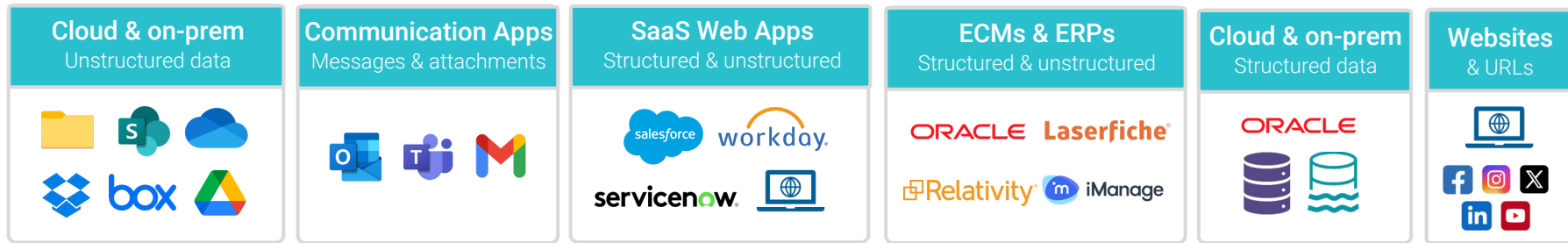
Where do DM and RIM/IG intersect?

How to Get Started


Data Governance Challenges



Data Governance Challenges



FIND
Locate & Identify



EVALUATE
Inventory & Analyze




TAG
Classify & Report



ACTION
Review & Disposition



ENSURE
Monitor & Audit



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

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

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

- Interactive & instant reporting
- Automated rules & workflows
- Defensible disposition

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

Data Governance Challenges

Cloud & on-prem
Unstructured data



Communication Apps
Messages & attachments



SaaS Web Apps
Structured & unstructured



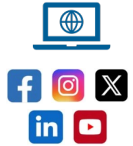
ECMs & ERPs
Structured & unstructured



Cloud & on-prem
Structured data



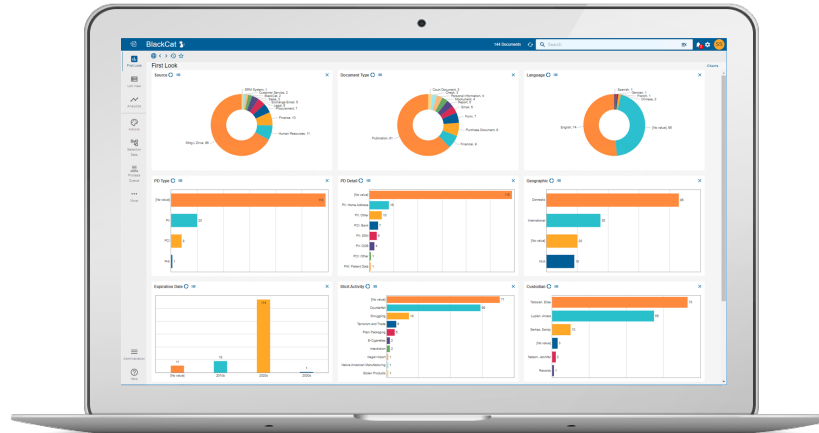
Websites
& URLs



- ✓ What do we have?
- ✓ Where is it?

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

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



True AutoClassification



+



+



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?

Sophisticated Classification



File & Rich Metadata
Attributes

+

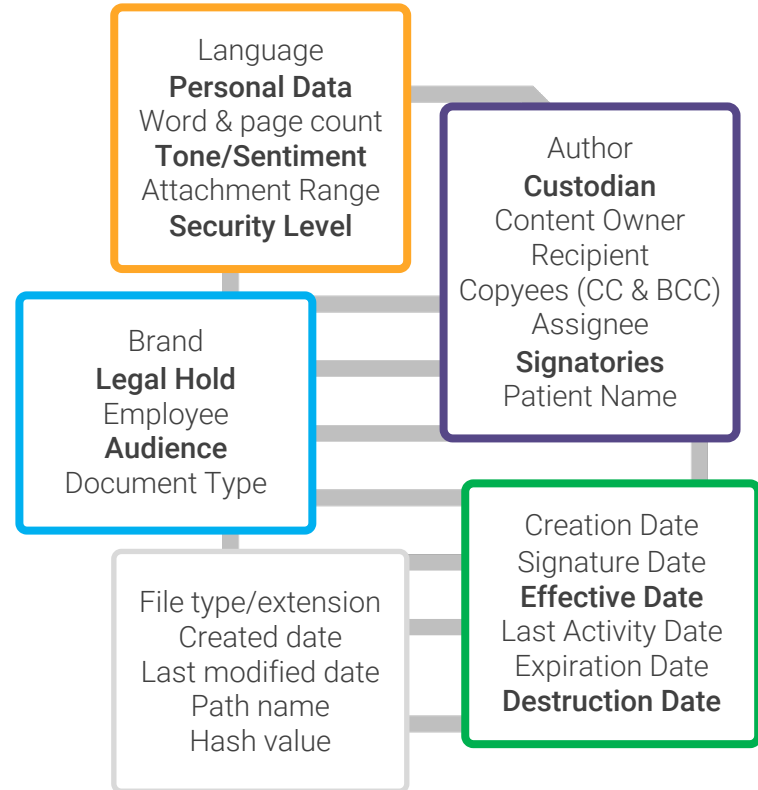


Recognition Algorithms
Determine DocType

+



Disposition Algorithms
Rules & Actions



Highly customized



File & Rich Metadata
Attributes

+

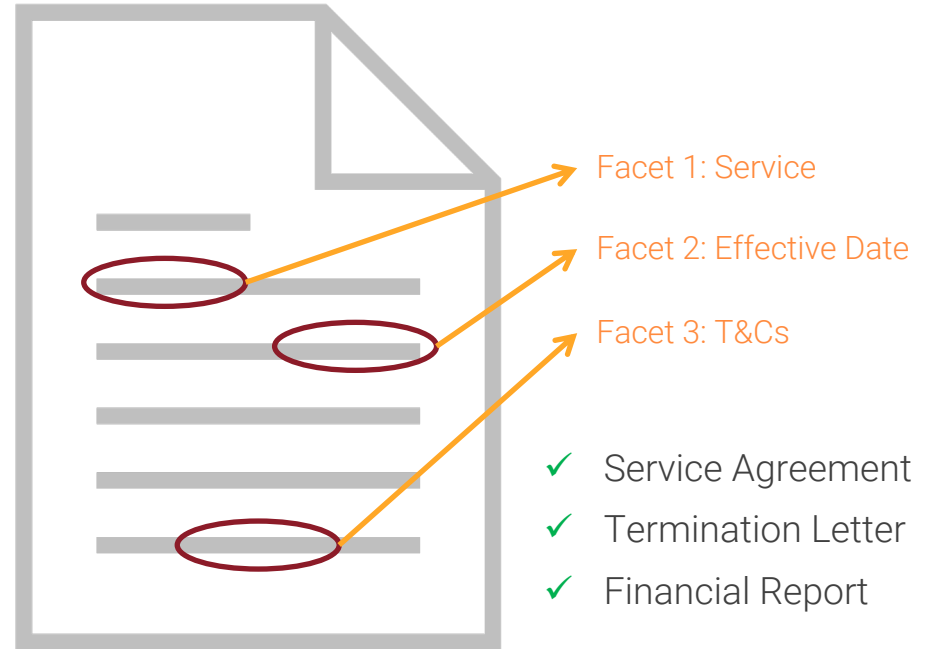


Recognition Algorithms
Determine DocType

+



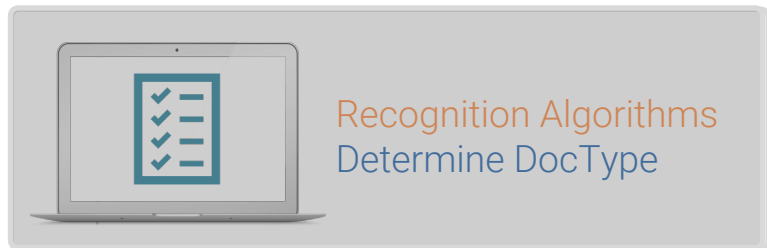
Disposition Algorithms
Rules & Actions



Comprehensive results



+



+



- Determine the disposition & handling of content
- Typically follow an IF-THEN format, often nested



Disposition Rule Example:
IF the *DocumentType* = Contract AND
the *Date* < 1/1/2027, **THEN**
Mark File as ACTIVE (Retain CY +7)

Enterprise-wide Data Governance



Classification



ROT



Retention



Legal Hold



eDiscovery



Data Privacy




Data Security



Migration

AutoClassification turns this...

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

Into this...



Local Government Property Insurance Fund Actuarial Service **Contract**

THIS AGREEMENT effective this **1st 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 1st day of February 2005, and expiring on the 31st 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

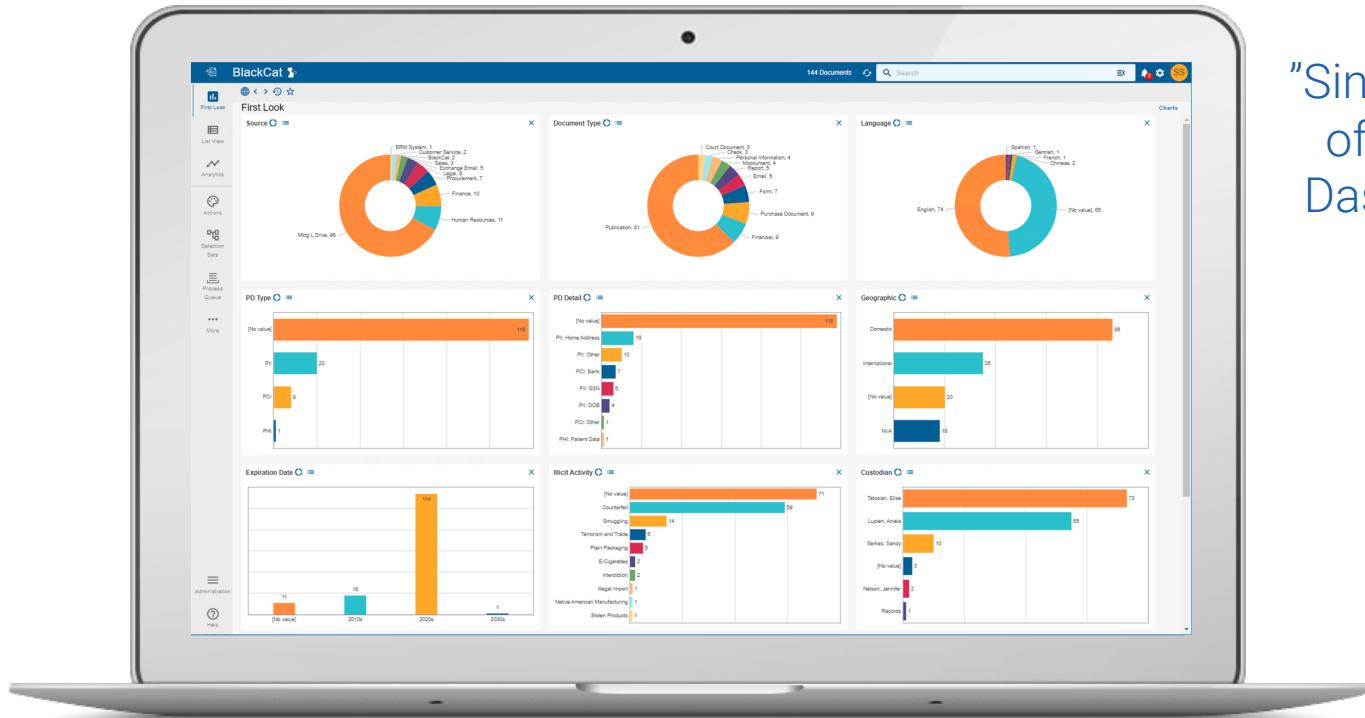
Handling = Delete after
authorization

Sensitivity = Confidential
(protection/redaction based on user
access class)



Retain or Delete?

Into this...



“Single Pane
of Glass”
Dashboard
View

Big 3 AutoClassification Use Cases

Data Governance



Reduces data to records of business value, rich metadata attribution & rules

- ROT Processing
- Rich Metadata Attribution, AutoClassification by DocType
- Records Management
- Workflow Automation
- System Migrations

Privacy & Compliance



ID files that contain personal data, mitigate risk, comply with regulations

- Privacy Regulations / DSAR
- Data Discovery
- Defensible Risk Mitigation
- Regulatory Compliance
- Reporting

Legal / Finance / HR



Reduces effort, reduces turnaround, defensible decisioning and audit trail

- Early Data Assessment: data set reduction / eDiscovery
- Legal Hold
- Virtual Vault contracts mgmt
- Mergers, Acquisitions & Divestitures: DD, Post-Merger Integration (PMI)

Agenda

Introduction to Valora

What is AutoClassification?

AutoClassification for Records Mgmt

What is Data Minimization?

Where do DM and RIM/IG intersect?

How to Get Started

Records & Information Management Use Cases



Identify & tag content with rich metadata

- Paper & electronic
- Multi-repository view
- Which files are records?
- Which files are ROT (junk)?
- Which files contain sensitive content?



Setup for disposition & lifecycle management

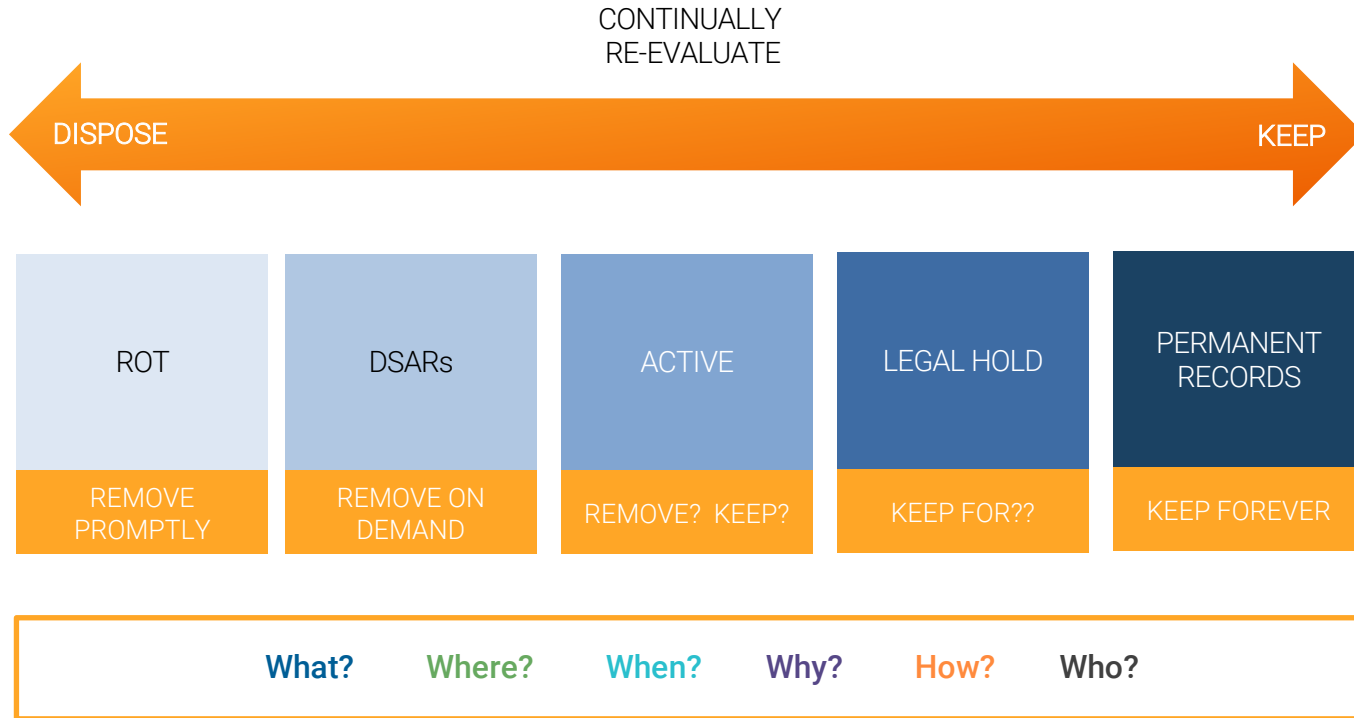
- Label Record Types & Classes for each file
- Enact proper retention –
- Dictate disposal method
- What authorizations?



Advanced Lifecycle Management

- Subordinating to Legal Hold
- Handling future states:
 - New/changed content, policies, regulations
 - New/changed objectives, strategies, budgets
- Assisting other groups in data management, retention and production (ex: Legal, KM, M&A)

Retention as a Spectrum



Typical ROT & Retention Classification Tags



ROT YES/NO

- Simple umbrella classification Yes/No
- Summary of ROT Detail



DUPLICATE TYPE

- Identical (Hash) Duplicate
- 99% (Functional) Duplicate
- 75-98% (Near) Duplicate
- Unique



RECORD CLASS

- Based on Document Type, Source, Custodian/Author, Jurisdiction, Keywords, Personal Data/Sensitivity



ROT DETAIL

- Redundant: Identical Dupe
- Obsolete: File Extension, Past Retention, Temporary, Draft
- Trivial: Auto-Reply, Low Business Value, Non-Record, Code or Executable, Thumbnail, Logo, File Path, File Type, Logo



RETENTION PERIOD

- Direct inherit from Record Class (Rules)



EXPIRATION DATE

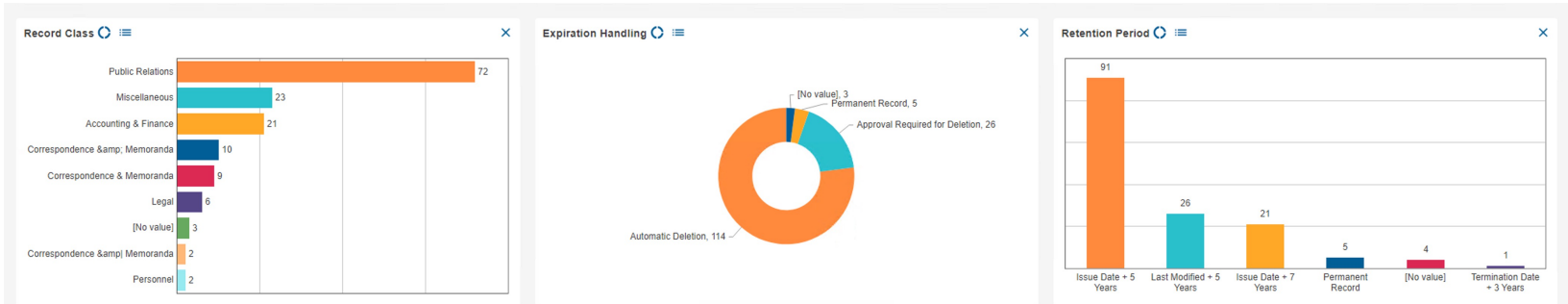
- Calculated from Date Hierarchy (differs by Records Type)
- Creation Date, Hire/Start Date, Termination Date,
- Often triggered by events



EXPIRATION HANDLING

- Automatic Deletion
- Approval Required for Deletion
- Quarantine
- Permanent Record

Automated Retention Handling



List View

Original File Name	Document Type	Record Class	Expiration Date	Expiration Handling	ROT	ROT Type	ROT Detail
HR_Job Application 1.jpg	Form	Miscellaneous	10/7/2021	Automatic Deletion	Yes	Obsolete	Obsolete: Past Retention
NeuroCure Clinical Trial Form...	Form	Miscellaneous		Automatic Deletion	No		
Meeting Materials1.jpg	Meeting Materi...	Legal	11/1/2021	Permanent Record	Yes	Obsolete	Obsolete: Past Retention
HR_Resume 1.png	Personal Infor...	Miscellaneous	10/10/2021	Automatic Deletion	Yes	Obsolete	Obsolete: Past Retention
HR_Resume 2.png	Personal Infor...	Miscellaneous	10/7/2021	Automatic Deletion	Yes	Obsolete	Obsolete: Past Retention
HR_Resume 3.png	Personal Infor...	Miscellaneous	10/10/2021	Automatic Deletion	Yes	Obsolete	Obsolete: Past Retention
Doctors warned about counterf...	Publication	Public Relations		Automatic Deletion	No		
ICE targets sale of NBA count...	Publication	Public Relations		Automatic Deletion	No		
The Scary Truth About the Bea...	Publication	Public Relations		Automatic Deletion	No		
Closing Statement.png	Purchase Doc...	Accounting & Finance		Approval Required for Deletion	No		
Closing Disclosure2.png	Purchase Doc...	Accounting & Finance		Approval Required for Deletion	No		
HR_Incident Report Form 1.jpg	Report	Miscellaneous	10/10/2021	Automatic Deletion	Yes	Obsolete	Obsolete: Past Retention
Banking Statements 2.jpg	Purchase Doc...	Accounting & Finance	5/31/2010	Approval Required for Deletion	Yes	Obsolete	Obsolete: Past Retention
HR_Job Application 2.gif	Form	Miscellaneous	10/7/2021	Automatic Deletion	Yes	Obsolete	Obsolete: Past Retention
HR_Timecard 1.jpg	Form	Miscellaneous	10/7/2021	Automatic Deletion	Yes	Obsolete	Obsolete: Past Retention
Personal Information3.jpg	Personal Infor...	Personnel	11/1/2021	Automatic Deletion	Yes	Obsolete	Obsolete: Past Retention
Suplies Expenses.gif	Financial	Accounting & Finance	1/1/2014	Approval Required for Deletion	Yes	Obsolete	Obsolete: Past Retention
Purchase Order1.jpg	Purchase Doc...	Accounting & Finance	1/9/2015	Approval Required for Deletion	Yes	Obsolete	Obsolete: Past Retention

How does AutoClassification help?



Identifies & Tags

Separates the ROT from content of business value



No Guesswork

Remove inconsistencies, drudgery of manual efforts



Cost Savings

Storage costs, human costs



Creates Instruction

at the content level, based on metadata & rules



Productivity

Maximizes human effort, increases productivity



Evergreen

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



Defensible

Actions, decisions, movement & access is logged



Set It & Forget It

RRS/policy updates, Record classes are all captured

Agenda

Introduction to Valora

What is AutoClassification?

AutoClassification for Records Mgmt

What is Data Minimization?

Where do DM and RIM/IG intersect?

How to Get Started

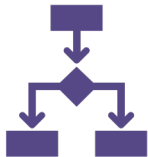
What is Data Minimization?



Data minimization, in the domain of Information Governance, is a principle and practice that involves **collecting, processing, and retaining only the minimum amount of data necessary** to achieve a specific purpose or business objective.



Compliance regulations require organizations to have **a lawful basis for processing personal data** and to ensure that they **do not collect or retain more data than is necessary** for the stated purpose.

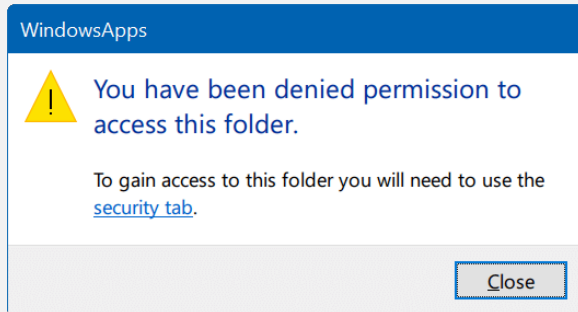


Data minimization is **a key component of responsible data management** and plays a crucial role in protecting individuals' privacy and reducing the potential for data breaches and misuse.

What is Least Privilege?

Current use of this term is typically in a **data security context**:

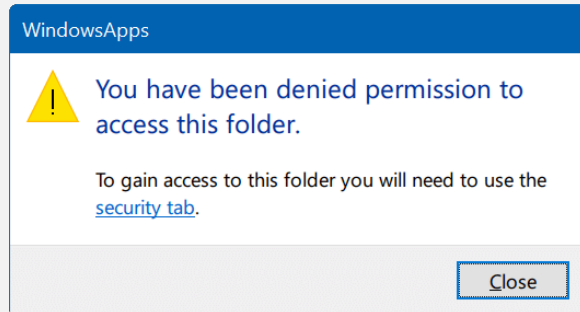
Least Privilege is a principle and access control strategy that restricts individuals (or systems) to only the minimum levels of access or permissions *to systems or applications* required to perform their job functions or tasks.



What is Least Privilege?

Current use of this term is typically in a **data security context**:

Least Privilege is a principle and access control strategy that restricts individuals (or systems) to only the minimum levels of access or permissions **to systems or applications** required to perform their job functions or tasks.



Emerging use of this term is in a **data governance/records context**:

Least Privilege is a principle and access control strategy that restricts individuals (or systems) to only the minimum levels of access or permissions **to data, records or content** required to perform their job functions or tasks.

39. On or about May 6, 2021, NARA made a request for the missing PRA records and continued to make requests until approximately late December 2021 when NARA was informed twelve boxes were found and ready for retrieval at the PREMISES. [REDACTED]

[REDACTED]

[REDACTED]

Data Minimization requirements are standard fare in data privacy regulations



Current Laws

- **GDPR:** any personal data collected (held) must be “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.”
- **CPRA:** business and organizations “shall not retain a consumer’s personal information or sensitive personal information . . . for longer than is reasonably necessary”

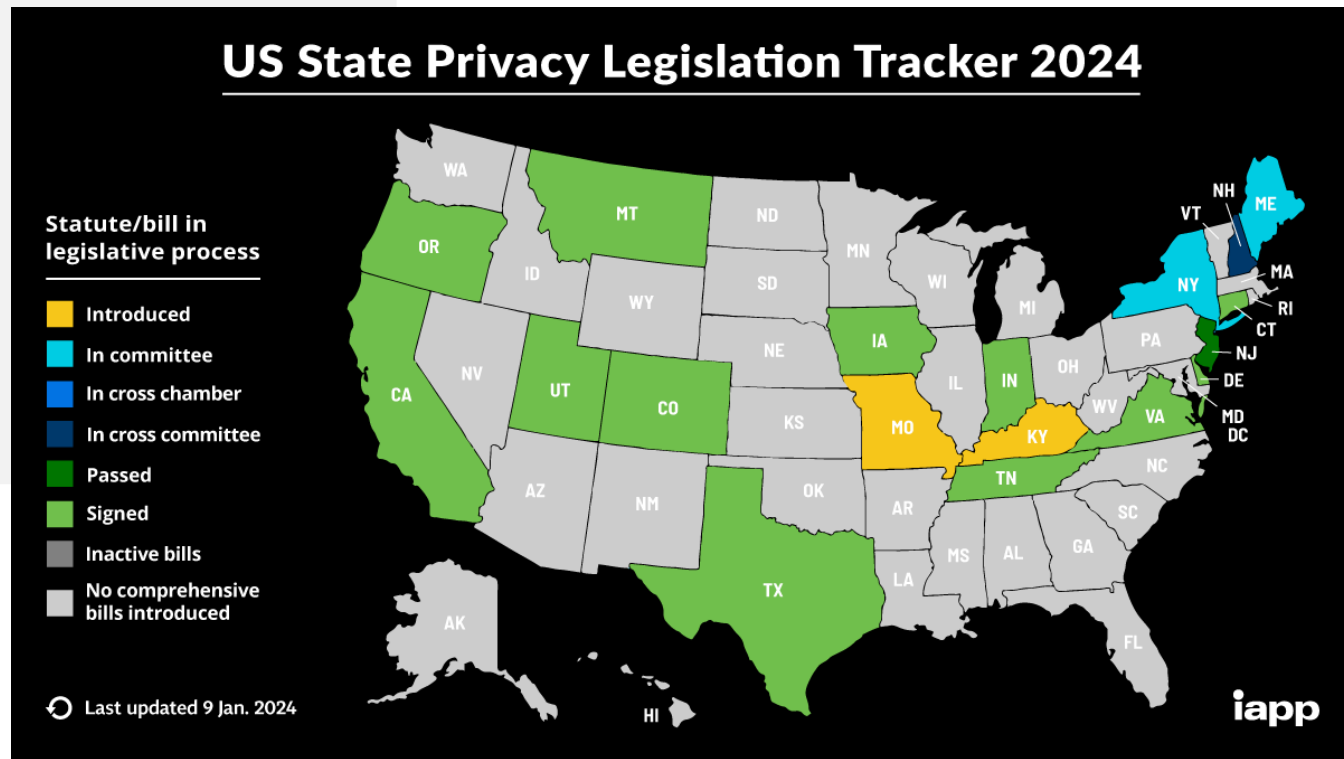


Still pending

- **Data Care Act:** “duty of confidentiality” and “duty of loyalty” that restrict disclosure or sale of personal data, particularly those that benefit the seller “to the detriment of the end user.”
- **Online Privacy Act:** “requirements for covered entities, including data minimization”

Why is this important now?

- Get to know IAPP
- Current status of data privacy regulations in the US (Jan 2024)



Why is this important now?

- Stop thinking data privacy is “someone else’s problem”
- Just like eDiscovery, this will become a RIM/IG “problem”
- Still time to get on top of it

Comprehensive Consumer Privacy Bills

State	Legislative process	Statute/bill	Common name	Consumer rights							Business obligations					
				Right to access	Right to correct	Right to delete	Right to opt out of certain processing	Right to portability	Right to opt out of sales	Right to opt in for sensitive data processing	Right against automated decision-making	Private right of action	Opt-in default (requirement age)	Notice/transparency requirement	Risk assessments	Prohibition on discrimination (exercising rights)
LAWS SIGNED (TO DATE)																
California		CCPA	California Consumer Privacy Act (2018; effective 1 Jan. 2020)	X	X		X	X			L	16	X		X	
		CPRa	California Privacy Rights Act (2020; fully operative 1 Jan. 2023)	X	X	X	S	X	X		X	L	16	X	X	X
Colorado		SB 190	Colorado Privacy Act (2021; effective 1 July 2023)	X	X	X	P	X	X	X	X-	S/13	X	X	X	X
Connecticut		SB 6	Connecticut Data Privacy Act (2022; effective 1 July 2023)	X	X	X	P	X	X	X	X-	S/13	X	X	X	X
Delaware		HB 154	Delaware Personal Data Privacy Act (2023; effective 1 Jan. 2025)	X	X	X	P	X	X	X	X	17	X	X	X	X
Indiana		SB 5	Indiana Consumer Data Protection Act (2023; effective 1 Jan. 2026)	X	X	X	P	X	X	X	X-	S/13	X	X	X	X
Iowa		SF 262	Iowa Consumer Data Protection Act (2023; effective 1 Jan. 2025)	X		X		X	X			S/13	X		X	X
Montana		SB 384	Montana Consumer Data Privacy Act (2023; effective 1 Oct. 2024)	X	X	X	P	X	X	X	X-	S/13	X	X	X	X
Oregon		SB 619	Oregon Consumer Privacy Act (2023; effective 1 July 2024)	X	X	X	P	X	X	X	X-	S/13	X	X	X	X
Tennessee		HB 1181	Tennessee Information Protection Act (2023; effective 1 July 2025)	X	X	X	P	X	X	X	X-	S/13	X	X	X	X
Texas		HB 4	Texas Data Privacy and Security Act (2023; effective 1 July 2024)	X	X	X	P	X	X	X	X-	S/13	X	X	X	X
Utah		SB 227	Utah Consumer Privacy Act (2022; effective 31 Dec. 2023)	X		X	P	X	X			13	X		X	

Who is leading the way?



Healthcare

Restricts access to patient data, only authorized personnel can view & use



Financial Services

Limits the collection and retention of financial information



Retail & eCommerce

Streamlines customer data management and protect payment card information



Technology & Software

Implements data minimization practices in their software and platforms



Govt & Public Sector

Implements data minimization practices in their software and platforms



Energy & Utilities

Minimizes the data collected from smart meters and IoT devices

OK, ok, data privacy, I get it. Is there any other reason or driver to get on board with data minimization?



Efficiency &
Cost Reduction



Performance
Optimization



Simplified Data
Governance



Reduced Legal &
Regulatory Risk



Environmental
Sustainability



Improved
Data Quality



Enhanced
Data Security



Improved
Decision-Making



Streamlined
Data Migration



Content Lifecycle
Management

... just a few!

Agenda

Introduction to Valora

What is AutoClassification?

AutoClassification for Records Mgmt

What is Data Minimization?

Where do DM and RIM/IG intersect?

How to Get Started

Organization's Data ~~Privacy~~ Responsibilities



DATA PRIVACY

What sensitive data do we have?

Where is sensitive data located?

Who has access to sensitive data?

Are we handling sensitive data correctly?

What sensitive data is accessible that shouldn't be?

Can I produce the relevant information when a DSAR request is made?



RECORDS MANAGEMENT

What information do we have?

Where is our data located?

Who has access to what data?

Are we handling our corporate data correctly?

What corporate data is accessible that shouldn't be?

Can I produce the relevant information when a request is made?

Which group or department is typically responsible for data minimization practices?

Like information governance, data minimization is a multidisciplinary effort, often involving:



1. Chief Information/Data Officer (CIO/CDO)
2. Data Governance Team
3. Data Privacy or Compliance Office
4. Information Security Team
5. Legal and Compliance Department
6. IT Department
7. Business Units and Data Owners
8. Data Stewards
9. Audit and Internal Controls Teams
10. Training and Awareness Teams

Describes you?
Sounds like RIM/IG?

What does a typical DM program/policy look like?

- It looks like a records retention schedule!
- With some key additions:
 - ✓ Least privilege – who can see/do what with this file/record
 - ✓ Intended lifecycle vs. record lifecycle
 - ✓ What time/event triggers force the disposition forward?
 - ✓ Data privacy citations
 - ✓ Sensitivity
- This means single-tab record-keeping is not enough
 - Record Class is insufficient (and non-compliant!)
- New retention/disposition labels required to retain only as long “as needed” (per the stated intention at collection/creation time)

Things can get complicated quickly:

We keep client transaction data for 3 years past the transaction date

- unless the client requests their “right to be forgotten,” in which case it is 45 days from DSAR date
- unless it is under legal hold.

Watch out!

Suddenly “past retention” takes on new, onerous meaning

- It's no longer ok to keep data “past retention” because it is in direct violation of data minimization that is expressed as “no longer than...”
- Vs records retention, which is implicitly expressed as “no shorter/less than...”
- New lifecycle management tags, expression, and management policy as a retention “window:”
 - Minimum retention period (RIM) – no less than
 - Maximum retention period (DM) – no more than
 - Subject to current Legal Hold

Remember:

If you retain it,
it is discoverable!

Historically, why were you unable to delete ROT?

- “No tools.” (Delete, delete, delete...)
- “Can’t find it, ROT is mixed in with the good stuff.”
- No one will let us delete things.



No real penalty to “just keep forever” (and violate our RRS, but ok)

Now, there are real consequences
to not deleting data that is supposed to be deleted (past retention), and that we
said we would delete/retain for only X purposes or time period.

We are now out of compliance (not just sloppy).

Meh, so we're out of compliance what's the big deal?

- GDPR fines for data privacy compliance violations
 - Meta - €1.2 billion (\$1.3 billion)
 - Amazon - €746 million (\$781 million)
 - Instagram - €405 million (\$427 million)
 - Facebook - €265 million (\$275 million)
 - WhatsApp - €225 million (\$247 million)
 - Google LLC - €90 million (\$99 million)
 - H&M - €35 million (\$41 million)
 - British Airways - €22 million (\$26 million)
 - Marriott International - €20.4 million (\$23.8 million)



How can tools like AutoClassification help establish, manage and maintain the principles of data minimization?

AutoClassification tools can play a significant role in establishing, managing, and maintaining the principles of data minimization within an organization. These tools use automated processes, machine learning, and predefined rules to categorize and classify data based on its content, context, and relevance.



Identification of Sensitive Data



Data Classification



Data Retention Policies



Access Control & Permissions



Data Minimization in Backup & Archives



Data Discovery & Mapping



Compliance with Privacy Regulations



Data Incident Response



Policy Enforcement



Regular Auditing & Monitoring

Agenda

Introduction to Valora

What is AutoClassification?

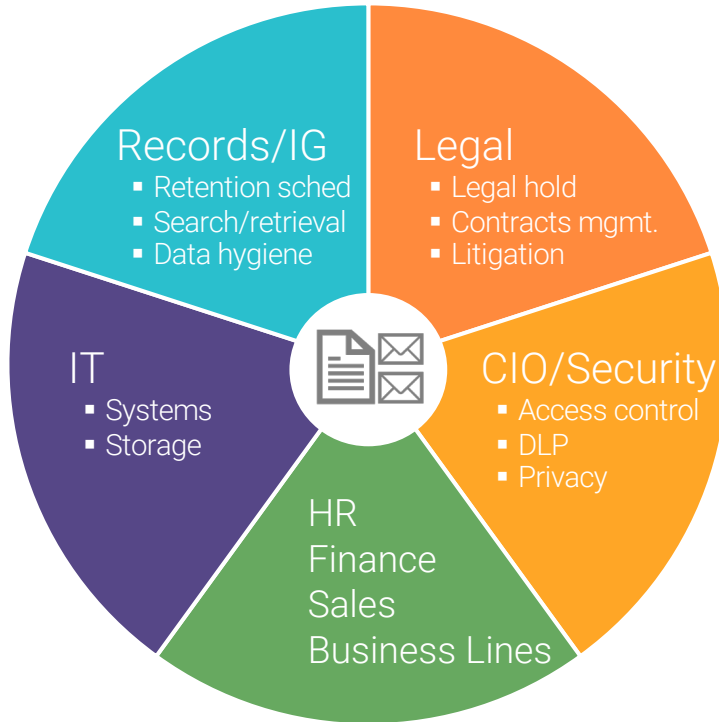
AutoClassification for Records Mgmt

What is Data Minimization?

Where do DM and RIM/IG intersect?

How to Get Started

Who will you need to get on board?

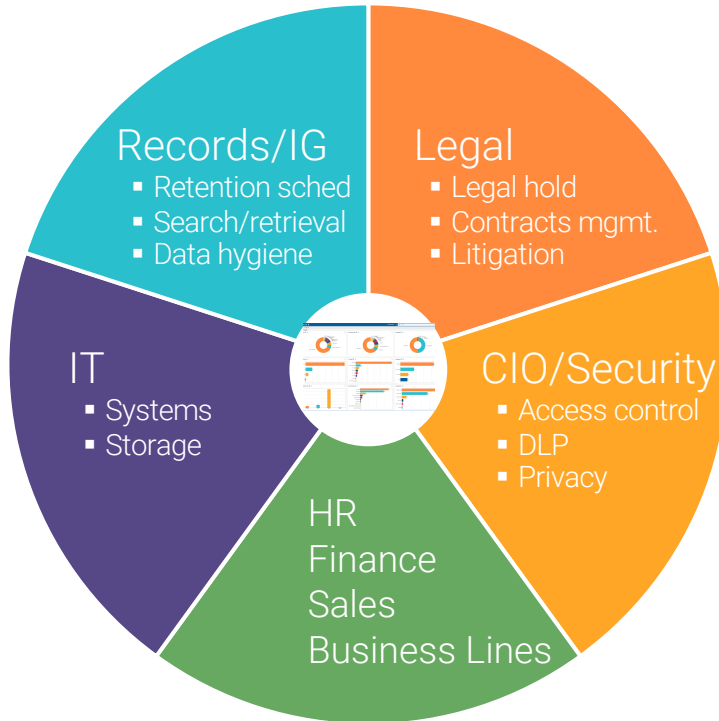


Every piece of content created has different value to each group

- ✓ 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

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

How to stand up a data minimization program in 10 “easy” steps



You will need:

1. An understanding of the regulations and legal basis for your requirements
2. Policies documenting how you will operate and comply
3. Executive support to implement, maintain and enhance these programs
4. Appropriate tools to execute and ensure compliance
5. Ongoing diligence and management



You will get:

1. Data inventory (aka discovery, mapping)
2. Establish tagging/retention criteria
3. Baseline tagging & disposition
4. Remove low hanging fruit (ROT, dupes, past retention, etc.)
5. Implement automated retention that is data privacy-aware (ensures data minimization)

How long does it take to set up or configure an AutoClassification tool to perform this work?



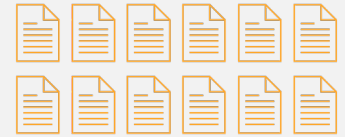
Complexity

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



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



Jennifer Nelson
VP Strategic Solutions
jnelson@valoratech.com
