



ORDER FROM CHAOS

Real World Lessons Using AI-Enhanced Auto-Classification

WELCOME

Jeffrey Sharer

Chief Executive Officer
LexShift

Jason Stearns

Chief Success Officer
LexShift



lexshift

GOVERNANCE REIMAGINED

Source:

INTRODUCTION

Jeffrey Sharer

- Chief Executive Officer, LexShift
- Former AmLaw 10 partner and founder of a national Data Law practice
- 20+ years advising on data, privacy, and regulatory risk
- Bridges legal strategy with practical technology execution

Jason Stearns

- Chief Success Officer, LexShift
- 20+ years leading Information Governance programs at global financial institutions
- FAI, IGP, CRM, CIPP/US, CIPM
- Focused on turning policy into operational, defensible execution

WHAT WE WILL COVER



Why classification is breaking



What AI changes



How it works in practice



Case studies



What we learned

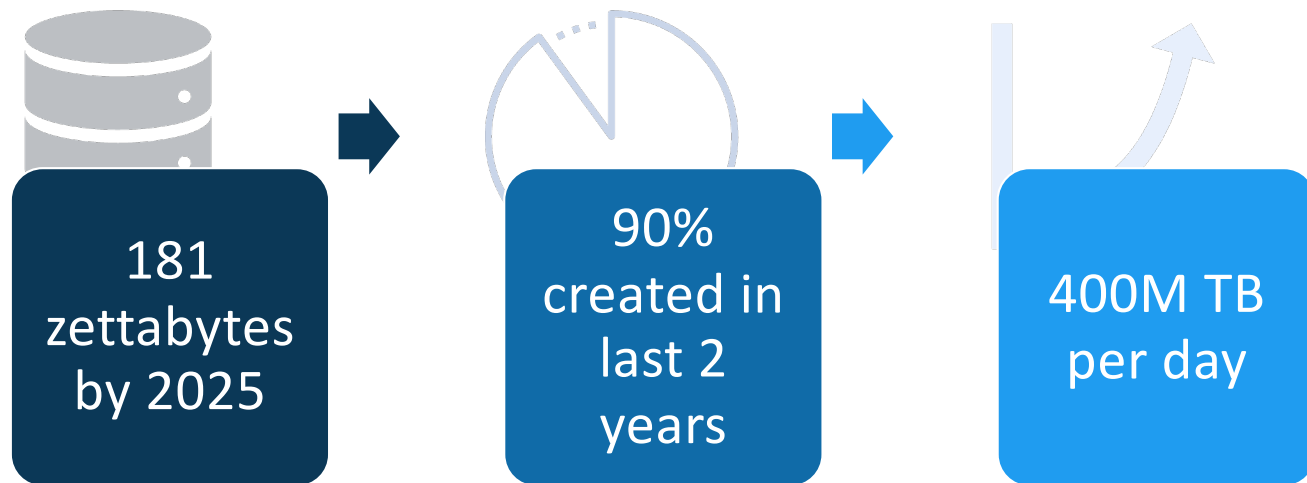


Where to start

THE REALITY

- Data is growing faster than governance
- Most organizations have policies, but don't have control
- Policy \neq Governance

HOW MUCH DATA?



WHY CLASSIFICATION IS BREAKING

The problem

Volume

Unstructured data

Inconsistent application

Why traditional methods fail

Manual tagging

User-driven classification

Static retention schedules

WHY CLASSIFICATION IS BREAKING

Policy defines
intent

Data reflects
reality

They are not
connected

WHAT AI-ENHANCED CLASSIFICATION ACTUALLY CHANGES

It is

- Pattern recognition across content
- Classification at scale
- Policy-aligned decisions

It is not

- Magic
- 100% accurate
- A replacement for governance

WHAT AI-ENHANCED CLASSIFICATION ACTUALLY CHANGES

Works well

- ROT identification
- Large-scale inventory
- Pattern-based classification

Struggles

- Ambiguous content
- Poor policy/retention schedule definition
- Lack of context

Critical requirement

- Show the work
- Transparency = Defensibility

8

AI NEEDS STRUCTURE. AND IT NEEDS HELP.

AI reflects your
model.
Good or Bad.

More categories =
more confusion

Ambiguity doesn't
disappear...
it gets amplified!

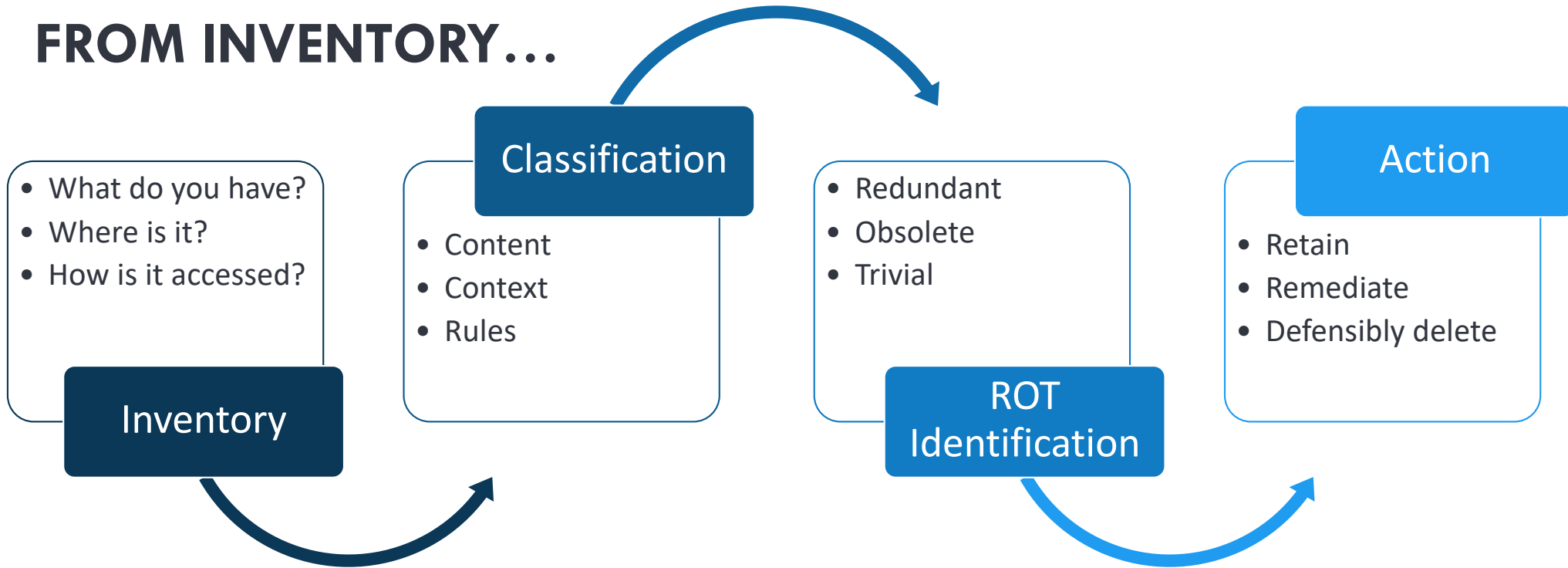
Humans-in-the-
loop are not
optional!

They define scope.
They remove
ambiguity.
They make it work.

This is how
accuracy becomes
defensible

HOW IT WORKS IN PRACTICE

FROM INVENTORY...



TO ACTION

10



**THE CASE STUDIES
THREE REAL-WORLD EXAMPLES. SAME PATTERN.**

CASE STUDY 1 – NON-PROFIT CORPORATION

The Client

- A private nonprofit corporation affiliated with a public university system

The Challenge

- Migrating content from legacy shared drives to SharePoint
- No record retention schedule and limited records management guidance
- No formal efforts to manage, retain, or delete content; inconsistent, department-driven practices
- Concerns about sensitive data being stored inappropriately

CASE STUDY 1 – NON-PROFIT CORPORATION

Approach

- Basic ROT Clean-up identification first, including identification of sensitive information
- Develop RRS in parallel department by department
- Delete information not being migrated
- Structure Teams/SharePoint to match RRS Classifications

What Changed

- Significant reduction in data volume
- Improved operational efficiency in migration
- Increased visibility into sensitive data and risk

CASE STUDY 1 – NON-PROFIT CORPORATION

Lessons learned

- Need to “box in” POC
- Start with a supportive, but tough audience
- Clearer expectations on RRS development timelines, challenges
- ROT remediation can precede RRS with controls
- Be flexible with running classifications before or after migration
- AI makes ongoing compliance with classification structure attainable and part of the normal course of business

CASE STUDY 2 – LIFE SCIENCES

The Client

- A global, research-driven life sciences organization operating in a highly regulated environment, managing large volumes of clinical, regulatory, and operational data across multiple systems.

The Challenge

- Retention Schedule and supporting policies exist
- Unstructured data has primarily been managed manually
- Files must be classified against the retention schedule AND 1000s of product names
- Content types are varied and are spread across multiple repositories

15

CASE STUDY 2 – LIFE SCIENCES

Approach

- Classify test data against full retention schedule
- Identify all applicable product names
- Refine classifications, rules, scope notes and repeat

What Changed

- Classification against retention schedule had +65% accuracy on first run
- Product names identification was highly accurate, but complicated by ‘families’
- Retention schedule enhanced with clearer descriptions, examples, exceptions, and other details

CASE STUDY 2 – LIFE SCIENCES

Approach

WHY ACCURACY STARTED LOW

- Refine classifications, rules, scope notes and repeat
- Classification scope included full RRS + 1000s of product names
- High complexity → high ambiguity → lower initial accuracy
- **AI behaved exactly like a human team would at this scale**
- Classification against retention schedule had +65% accuracy on first run
- Product name identification was highly accurate, but complicated by families'
- Retention schedule enhanced with clearer descriptions, examples, exceptions, and other details

They didn't have a classification problem. They had a scoping problem!

CASE STUDY 2 – LIFE SCIENCES

Approach

WHAT IMPROVED ACCURACY

- Refine classifications, rules, scope notes and repeat
- Iterative refinement of rules and scope
- Better definition of categories and examples
- **Active human-in-the-loop validation and tuning**
- Classification against retention schedule had +65% accuracy on first run
- Product names identification was highly accurate, but complicated by ‘families’
- Retention schedule enhanced with clearer descriptions, examples, exceptions, and other details

CASE STUDY 2 – LIFE SCIENCES

Lessons Learned

- Use cleaner datasets for initial runs
- Limiting classification rules can increase performance and accuracy
- If you ask AI to do too much, it will perform like a confused human
- Accuracy improves when you reduce scope, not when you add complexity
- “Out-of-the-box” results are the starting point, not the outcome
- Human-in-the-loop iteration is what makes results defensible

CASE STUDY 3 – ENERGY

The Client

- A Canadian offshore energy regulator responsible for overseeing exploration and production activities, ensuring compliance with safety, environmental, and operational requirements.

The Challenge

- 15+ year legacy repository in OpenText Documentum
- Migration to Microsoft 365 (SharePoint + OneDrive)
- Hybrid environment required (shared drives retained for large files)
- Decision to lift and shift all content, not selectively migrate
- Need to reclassify and clean data during migration

CASE STUDY 3 – ENERGY

Approach

- Lift-and-shift migration into SharePoint
- Apply AI-enhanced classification post-migration
- Use Illuminate to classify content against retention schedule AND detect and remediate ROT
- Manage-in-place across multiple repositories (SharePoint, OneDrive, Legacy shared drives)

What Changed

- Enabled cleanup of 15+ years of legacy content during migration
- Reduced migration complexity and cost
- Improved visibility into: Data inventory, Risk exposure, and Retention alignment
- Shifted from repository-dependent governance → policy-driven governance across environments

CASE STUDY 3 – ENERGY

Lessons Learned

- Without AI-enabled classification, this doesn't happen at this scale
- If you treat migration as a lift-and-shift, you're just moving the problem
- Pre-migration cleanup sounds right, but it usually fails at scale
- Classification after migration works because scope is simpler and better defined
- Legacy systems don't just store data, they accumulate risk
- There is no single-system solution

CASE STUDY – CROSS-CASE PATTERNS

What Worked

- Start with a focused POC
- Clarity is more important than AI sophistication
- Iteration > perfection
- Tight scoping of classification objectives
- Human-in-the-loop refinement cycles
- Well-defined classification scope (not just clear rules)

Where It Broke

- Access is complicated
- Don't review too much data during POC/Pilot due to diminishing returns
- Allocate dedicated time and staff to the POC/Pilot
- Unclear rules or lack of context leads to poor outcomes
- Overly complex classification models (too many categories, too much ambiguity)
- Expectation that AI would work without design and iteration

PRACTICAL GUIDANCE: WHERE TO START

Use AI

- High volumes of unstructured data
- Known ROT problem
- Regulatory pressures

Don't use AI

- No retention schedule
- No governance ownership
- Expectation of “set it and forget it”

PRACTICAL GUIDANCE: WHERE TO START

Requirements

- Defined policy
- Modern, robust retention schedule
- Data access
- Governance alignment

Common Mistakes

- Over-scoping
- Ignoring explainability and transparency
- Treating AI as the solution instead of an enabler
- Treating AI as plug-and-play instead of an iterative process
- Underestimating the need for expertise that understands BOTH IG and AI

WHY SOME AI CLASSIFICATIONS SUCCEED... AND OTHERS DON'T



High Performance

- Focused scope
- Clear definitions and examples
- Iterative refinement
- Human-in-the-loop



Low Performance

- 100+ categories
- Ambiguous definitions
- No iteration
- "Set it and forget it"

WHY SOME AI CLASSIFICATIONS SUCCEED... AND OTHERS DON'T

High Performance

- Focused scope
- Clear definitions and examples
- Iterative refinement

AI doesn't fail.

Poorly scoped problems do.

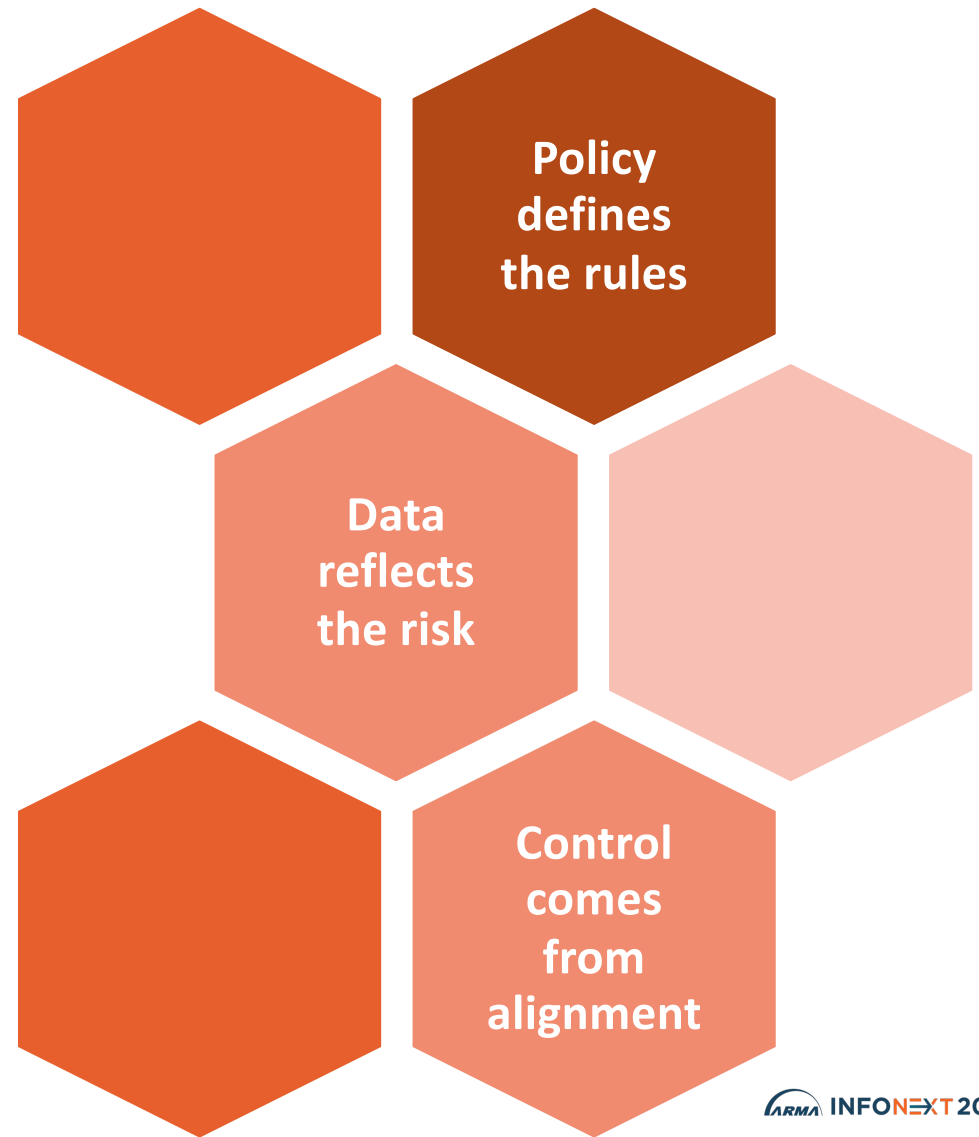
Low Performance

- 100+ categories
- Ambiguous definitions
- No iteration
- "Set it and forget it"

HOW THIS GETS IMPLEMENTED – OVERVIEW OF THE AI TOOL



FINAL THOUGHTS



QUESTIONS?

Jeffrey Sharer

Chief Executive Officer
LexShift

Jason Stearns

Chief Success Officer
LexShift

Connect with us!

