



From Cleanup to Catalyst

Evolving Your IG Operation
into a Scalable and
Defensible Firm Asset



CHRISTIAN PASCHKE
Product Advisor – Data Discovery
ActiveNav

- 6 Years information governance experience within FedEx legal department
- Implemented solutions for defensibly addressing structured and unstructured data across complex business and legal environment
- ActiveNav voice-of-the-customer advisor to product development and implementation teams



MARK EVANS
Director of Solutions Engineering
ActiveNav

- 20+ Years information management experience
- Completed world's first digital archive for UK National Archives and launched Preservica's North American digital archive platform
- ActiveNav Lead solutions architect for designing unstructured data discovery approaches at scale for legal market

- Information Governance Cleanup 2.0 for Business Value
- Programmatic Unstructured Data Discovery Methodology
- Scaling Smarter with Technology for Competitive Advantage





IG Cleanup 2.0 for Business Value



Unstructured Data Cleanup 1.0

- Focus on ROT (redundant, obsolete, trivial) data
- Emphasis on retention/deletion
- Project-oriented (shutting down the shared drive)
- Resource-intensive
- Results tied to reduction

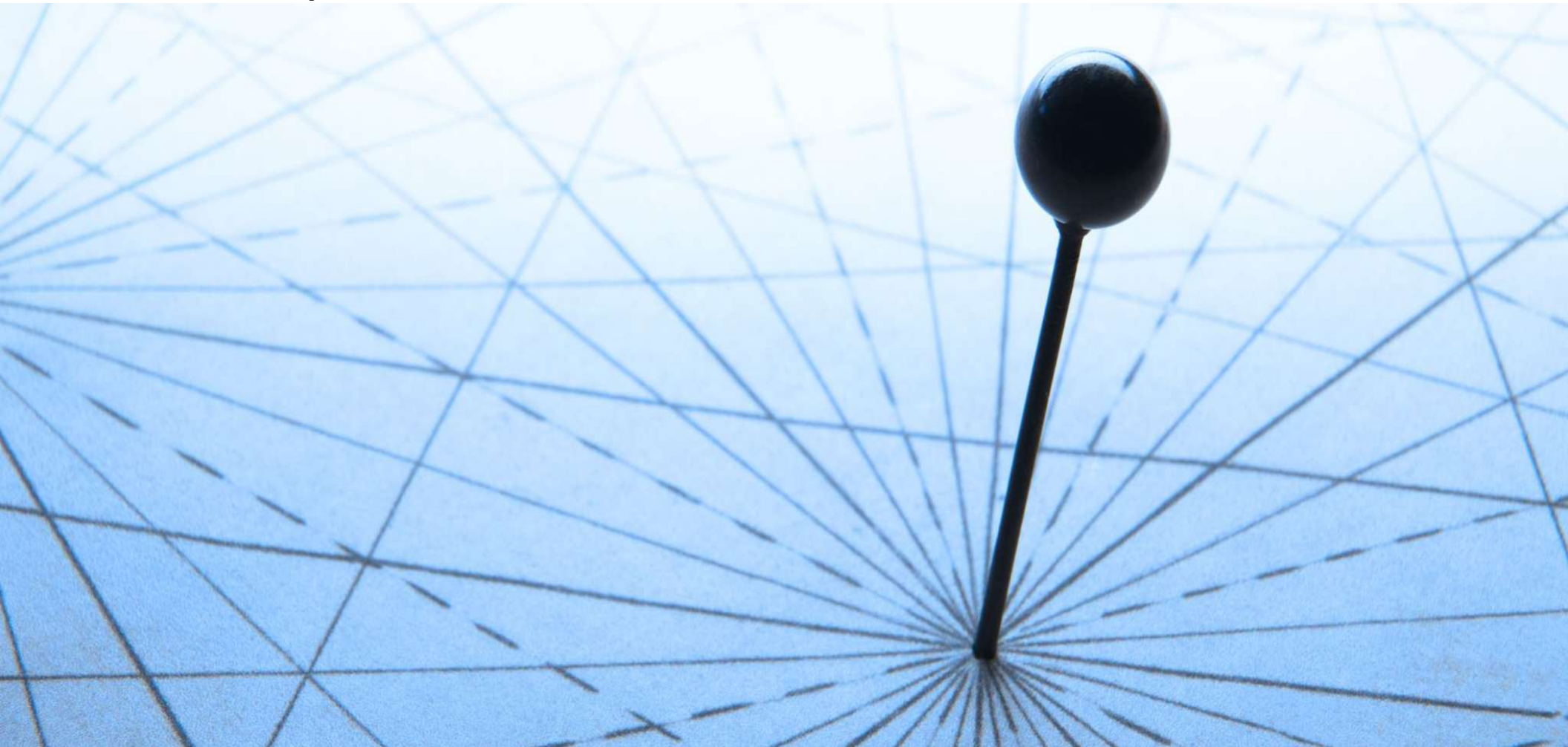


- Focus on business goals
 - Lifecycle content management
 - Cloud cost reduction
 - AI readiness and compliance
 - Knowledge management for competitive advantage
 - Data privacy and compliance
- Emphasis on data value
- Programmatic approach
- Technology enabled
- Results tied to organizational metrics





Programmatic Unstructured Data Discovery Methodology



Unstructured Data Management Methodology

EDRM IG Model

INFORMATION GOVERNANCE REFERENCE MODEL (IGRM)
Balancing Value, Risk and Cost



LFIGS IG Processes



Information Governance
Program Governance and Goals

Analyze & Validate
Review Results and Validate to Goals



Decide
Review Results and Determine Action



Act
Per Policy, Procedure, or Workflows



Sustain
Ongoing Refresh, Reporting, and Review



Policy & Rules
Procedures and Standards for Defensible Action

Data Topology
Modeling and Grouping, Orchestrating Discovery Process, and Organizing Reporting

Metadata Discovery
Inventory of Structure and Properties



Content Discovery
Inventory of Relevant Data from Content



Classification
Grouping, Clustering, and Organizing by Risk and Value






Inventory
Discovered Data with Key Properties and Metrics

Steering & Oversight
Establishing and Assessing Program Metrics to Make Data-Driven Decisions



Unstructured Data Management: Approach Profiles

APPROACH	 SPEED	 BALANCE	 PRECISION
USE CASES	<ul style="list-style-type: none"> • Cleanup: "Big Bucket" <ul style="list-style-type: none"> ▪ Redundant & Trivial ▪ Age-Based ▪ Individuals 	<ul style="list-style-type: none"> • Cleanup: Content-Driven <ul style="list-style-type: none"> ▪ Outdated ▪ Duplicates • AI Readiness • Privacy / Sensitivity • DSARs • Cybersecurity • Matter Mobility: Bulk & Container Review 	<ul style="list-style-type: none"> • Cleanup: Review-Driven <ul style="list-style-type: none"> ▪ Legal Hold ▪ Records Management Compliance • Matter Mobility: Full Content Review • Client Audit
GOALS	<ul style="list-style-type: none"> • Reduce Storage Costs • Increase Storage Space • Data Minimization 	<ul style="list-style-type: none"> • Policy Compliance • OCG Compliance 	<ul style="list-style-type: none"> • Legal Hold Preservation Compliance • Records Management / Retention Compliance
FOUNDATIONS	<ul style="list-style-type: none"> • Confidence in Legal Hold Preservation System • "Bright Line" Deletion Standards • ROT Rules (Uncontestable) 	<ul style="list-style-type: none"> • IG / RIM Policy • ROT Rules (Contestable) • Privacy Policy • Lifecycle Matter Management Process 	<ul style="list-style-type: none"> • Defensible Deletion Documentation Requirements • Retention Schedule • Legal Hold Procedure
RISK ASSESSMENT	<ul style="list-style-type: none"> • Confident 	<ul style="list-style-type: none"> • Validate 	<ul style="list-style-type: none"> • Verify Everything
TIMEFRAME	<ul style="list-style-type: none"> • Short 	<ul style="list-style-type: none"> • Moderate 	<ul style="list-style-type: none"> • Extensive



Scaling Smarter with Technology for Competitive Advantage



Top 10 Law Firm Reclaims Unfiled Matter Data

Problem:

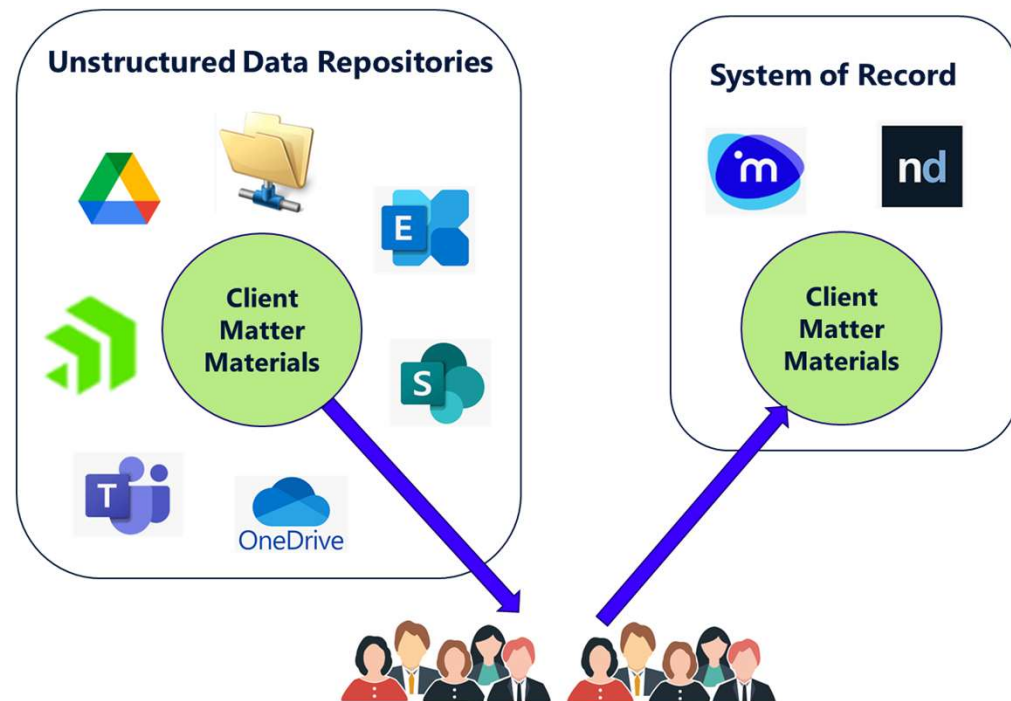
- Client matter materials stored outside the system of record

Scope:

- Initial target of 200k files to match against 1.2m matters

Challenges:

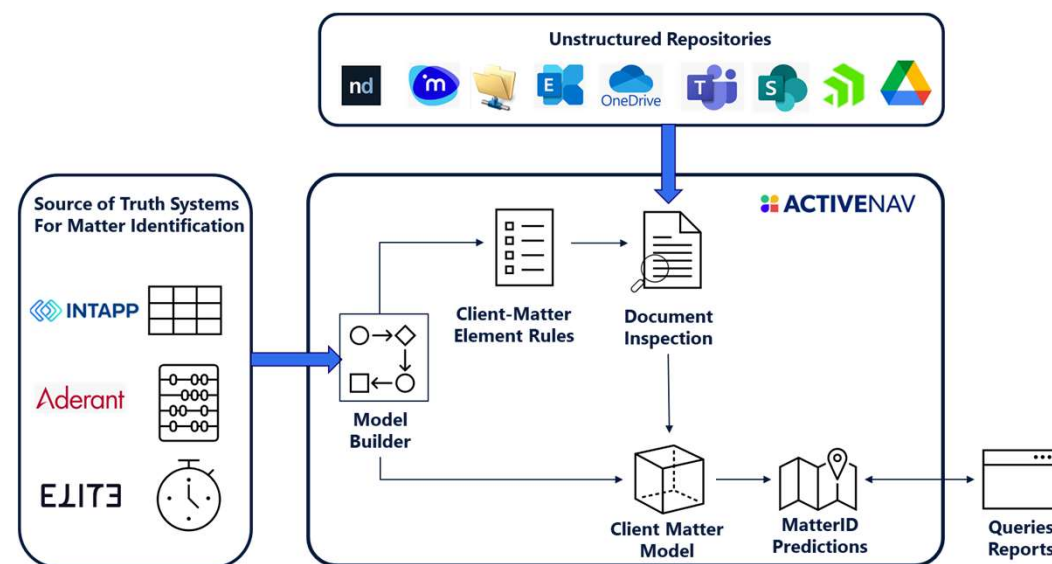
- Scale – Identifying and classifying files against thousands of possible matters
- Constraints – Traditional Search and eDiscovery methods won't work



Top 10 Law Firm Reclaims Unfiled Matter Data

Solution:

- Discovery criteria based on Source-of-Truth metadata sources
 - e.g. Intake and billing systems
- Custom model and rules to support precise identification
- Automated metadata and content analysis performed against model and rules for each file
- Confidence based matter prediction performed for each file
- Reports indicate responsive files for each matter within each repository



Top 10 Law Firm Reclaims Unfiled Matter Data



Analyzing and Validating Results at Scale



Matched object Confidence Score ranges fall into three buckets

Positive: Results with high confidence threshold, subdivided into

- Simple (obvious) predictions
- Complex predictions

Ambiguous: Confidence range not high enough to count as Positive, or multiple matches

- Complex predictions

Negative: Not considered a match



Performance measure assessment to determine ROI (technology vs human review)

Precision for the Positive set (P_p)

Precision for the Ambiguous set (P_a)

Recall (R) – The number of objects returned in both sets above sets, as a percentage of all possible matches in the entire discovered dataset



Performance Measure

Performance =
 $f(P_p, P_a, R)$

Top 10 Law Firm Reclaims Unfiled Matter Data

Statistical Sampling to Validate Results

Calculate sampling of all discovered objects

- 95% confidence and 5% margin of error
- Select samples across Positive and Ambiguous categories for human review and validation
- Calculate Performance Measure:
 - Performance = $f(P_p, P_a, R)$



- Aligning unstructured data discovery to **business goals** unlocks new value from IG teams
- Following a **repeatable methodology** for unstructured data discovery, classification, and action operationalizes information governance strategy
- **Technology assisted discovery** and classification generate defensible value at scale
- **Statistical sampling** with human validation enables defensible results and efficient ROI calculation



Questions?



CHRISTIAN PASCHKE
Product Advisor – Data Discovery
ActiveNav

christian.paschke@activenav.com
[Linkedin.com/in/christianpaschke](https://www.linkedin.com/in/christianpaschke)



MARK EVANS
Director of Solutions Engineering
ActiveNav

mark.evans@activenav.com
[Linkedin.com/in/mark-evans-39b0ab2](https://www.linkedin.com/in/mark-evans-39b0ab2)



www.activenav.com



Thank you for attending!

Twitter: @InfoGovWorld

LinkedIn: @InfoGov World Magazine

www.InfoGovWorldConference.com

AIWorldconference.ai