

Unstructured Data Wars

In a Galaxy closer than you think, the final frontier to manage all information

Contents

- System Components
- Getting the Word Out
- User Action Items
- Living the Dream
- Set Up the Environment
- Training
- Metrics
- Audit
- Learned Lessons

System Components

- Analysis Tool
- Knowledge Base
- Ingestion Tool
- Repository
- Repository User interface
- Repository Information Management
- Retention Tool

Analysis Tool

Pros

- Analyze file metadata
- Type
- Date Created
- Date Modified
- Duplicates
- Can locate PII and PCI

Cons

- Auto classify needs KB
- Event Based Retention
- PII PCI Need New Tool
- Firewall and Server Changes

Knowledge Base Tool

Pros

- Auto Classify
- Adjust Thresholds
- Significant Concepts
- Keyword Data

Cons

- Balance Samples
- Structured Forms
- Time consuming
- Cross Code Examples
- OCR PDFs

Ingestion Tool

Pros

- Automated ingestion
- Auto Metadata tagging
- Stubs
- Folder structure must be exact

Cons

- Task routes for all
- Tech Expertise
- Folder names must be exact
- Firewall changes
- Network Errors

Repository

Pros

- Settings comprehensive
- Interacts w AD
- Files are immutable

Cons

- Settings are complex
- Tech Expertise
- Changes can have a domino effect

Repository User interface

Pros

- Required Metadata
- Only Records
- Custom Desktops
- Custom Templates
- Master Template
- User managed custom metadata

Cons

- Loses metadata on non office doc types
- Templates = LOB (many)
- User managed custom metadata

Repository Information Management interface

Pros

- Applies Retention
- Legal Holds
- Approval Process

Cons

- Reporting

Getting the Word Out

- Engaging the LOBs
- Surveys - what they think they have
- Analyze – does it match?
- Kickoff meetings
- Customized Trainings

User Action Items

- Folder structure
- Required Metadata
- Metadata structure
- Once set up, management process is automated

Living the Dream

- Tested our theory on our own drive
- See how the system works with files we know
- Live what we are asking other groups to do
- System works
- Still a lag in user manual moving of the files

Set Up of Environment

- Access Security
- Access controls assist security & scoping
- Naming conventions - unique identifier
- User Facing Terminology
- Show Info for User Recognition
- Consider Naming Convention if info changes

Training

- Terminology for users – say what it takes so they can do what the system needs
 - These are the Company's Records
 - This is what the Company wants us to do with them
 - Following policy
 - Achieving Compliance
- Active vs Inactive - They don't have to understand RIM
- LOB to Understand Metadata Mapping for searches

Metrics

- Look for project justification
- We capture:
 - Size of Drive
 - # of Files
 - Ingestion Time

Audit

- Successful Processes Have Controls
 - Work w Audit what to look for
 - System controls / reporting
 - Spot Check
 - Communicate / Escalate

Learned Lessons

- Empty Folder Templates
- Test everything – don't assume
- Retention Trigger Not Met
- Outdated Systems
- Forms Control
- Metadata Issues – Office vs Google

Thanks

WTBASINGER@GMAIL.COM