Intersection of Information Management & Information Security

March 21, 2018

DOCULABS

DOCULABS

Management Consulting

Objective

Information Governance and Security

About Me

Matt McClelland is a Principal Consultant focused on Enterprise Information Management and Sensitive Data Identification. Matt has over 15 years of experience in Big Data Analytics and Information Management. Matt works with clients to conduct content and access controls scans of business-critical repositories, and develops client-specific data disposition plans for legally defensible cleanup and migration of business content.

- 919-623-3663
- mmclelland@doculabs.com
- www.doculabs.com

Agenda

- What’s the problem?
- How do I solve it?
- What’s in it for me?
- Bring it all together
- Q&A

Agenda

- What's the problem?
- How do I solve it?
- What's in it for me?
- Bring it all together
- Q&A

What's the Problem?

InfoSec has historically been focused on building stronger walls to keep bad actors out

But the state of information behind the walls is just as important

Terabytes of ROT and sensitive data (PHI, PII, PCI, IP) – most of it past its legal and operational life – are at risk for a breach

InfoSec must evolve to manage information with Records Management

What do we have?

Most organizations are not able to identify where PHI and PII are currently housed and are unable to identify the appropriate policies and procedures to take action against that data.

50%

Have a Data Map

65%

Are *not* Mapped to InfoSec Policy

75%

Are *not* Mapped to Records Policy

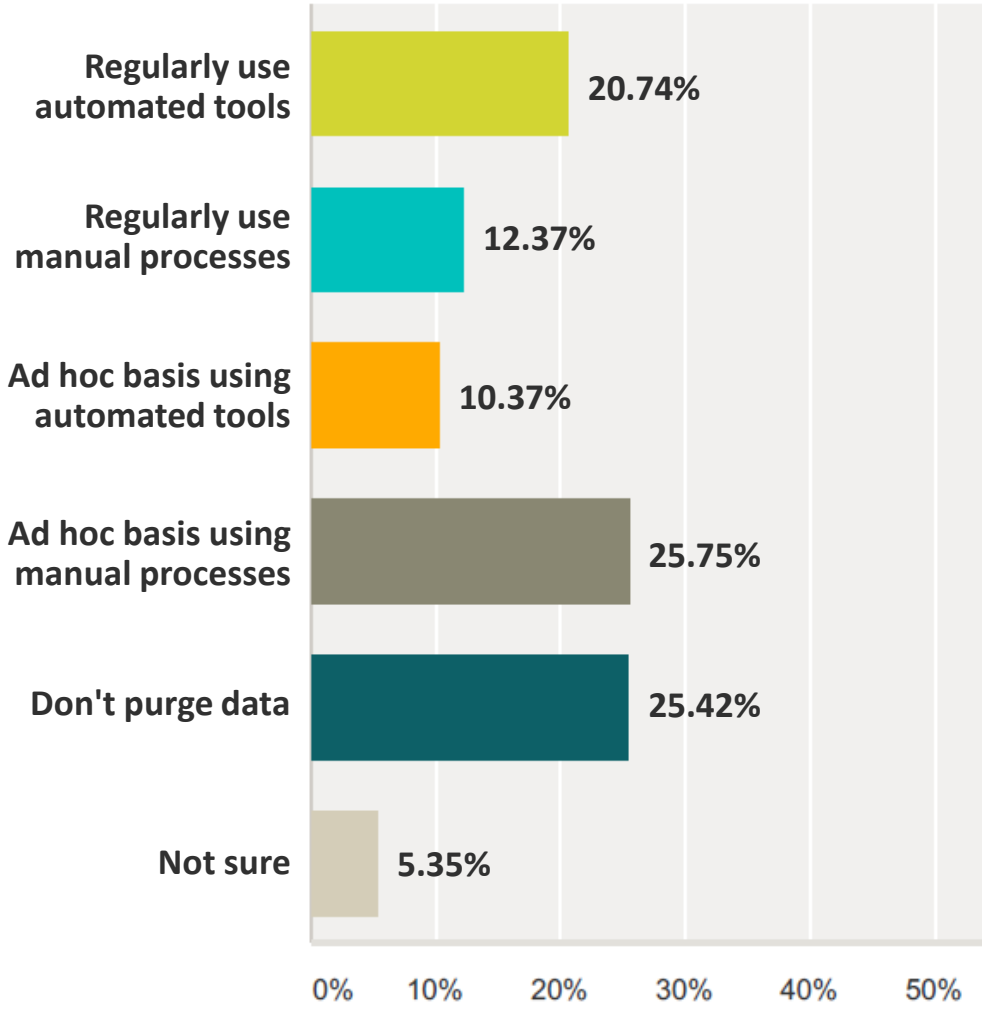
80%

Do *not* have an Orphaned Data Policy

When We Asked . . .

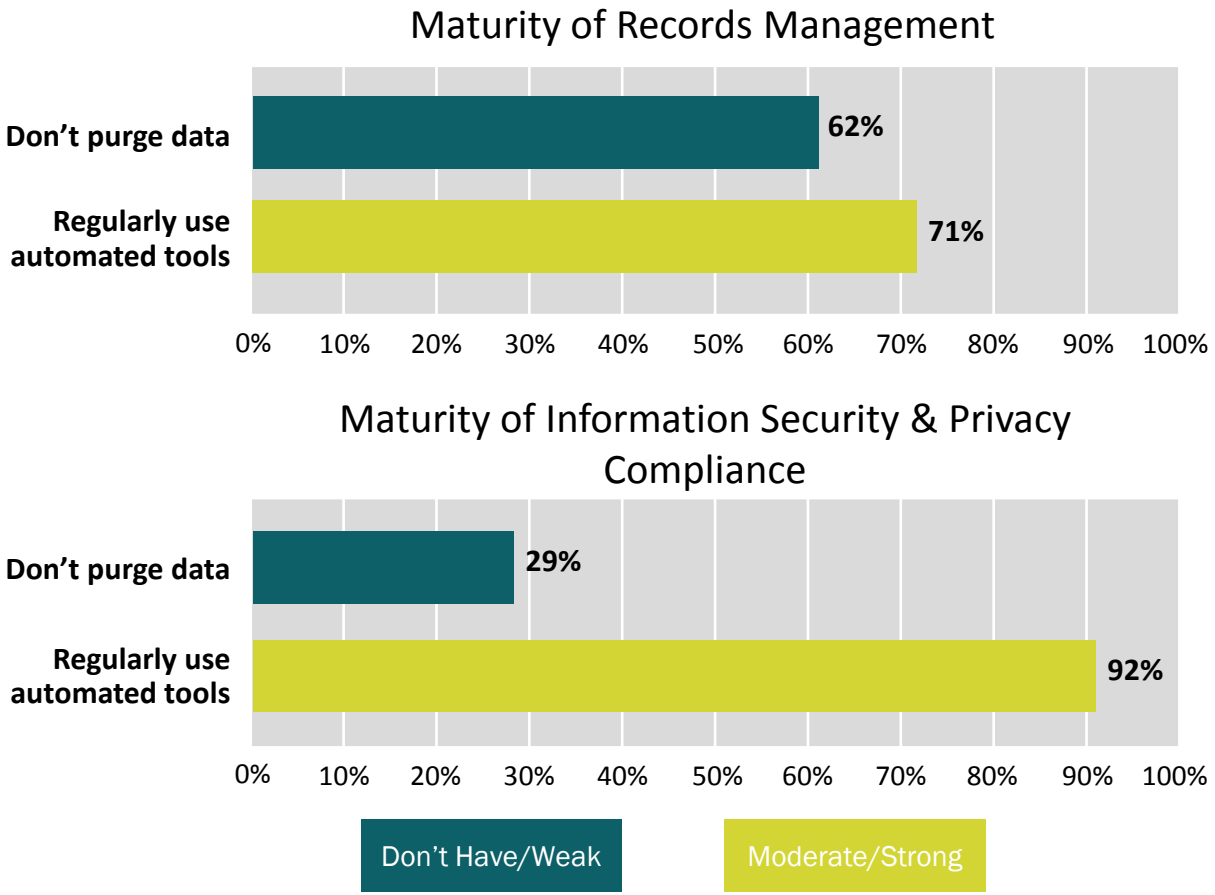
How

do you purge data after it has passed its legal or operational lifecycle?



Good governance is strongly correlated with regular data purging

Approach to data purging



- Firms that reported that they didn't purge data also reported that they had weaker records management, less mature information security and privacy, and less mature disaster recovery/business continuity functions.
- There was also a much stronger correlation between good data hygiene and good governance than between good data hygiene and technology capabilities.

Example Client: Shared Drive Repository Analysis

Shared Drives

Most commonly used managed repository for business user collaboration and convenience copies of content.

(Note there are approximately 2M files (10TB) that do not have a last accessed date available.)

Files	Capacity Used	Avg. Growth Rate (files) 2006-Present	Avg. Growth Rate (size) 2006-Present
72,100,000+	66 TB	65%	75%

Content Aging (last accessed)			
1-3 years	3-5 years	6-10 years	10+ Years
34.5M	23.7M	11.6M	139.5K
20TB	26TB	9TB	.023TB

Areas of Risk			
Folders with global access	Sensitive files	Sensitive files with global access	Aged sensitive files
Unknown	Unknown	Unknown	Unknown

File Type Breakdown					
Word	Excel	PDF	Access DB	PPT	Junk, Low Value, Other
6.4M	6.9M	6.5M	7.3K	413K	51M

How is it manifested?

How is it manifested?

Access

- Overly permissive access to sensitive data/IP and overly restrictive access to other data.
- As content grows, it becomes increasingly difficult for employees to find the content they need, as they sift through all the content they don't need.

Cost

- Content growth can be as much as 20 percent year over year, leading to increased data storage costs (including backups) over time, despite the general decrease in storage costs.
- Organizations face the high cost of outside counsel review of terabytes of irrelevant data during eDiscovery.

Effort

- Legal and eDiscovery efforts are significantly increased when content is over-retained.
- Information Security has a harder time protecting critical information when mixed in with junk and stale content.

Risk

- Keeping more sensitive data than required increases the severity of a breach (higher fines, more significant PR fallout).
- Non-compliance with corporate policies makes it more difficult to defend corporate conduct with regulators and the public.
- Cyber-security insurance may not pay claims in the case of non-compliance with organizational policies or specific insurance policies.

Agenda

- What's the problem?
- How do I solve it?
- What's in it for me?
- Bring it all together
- Q&A

Methodology for Content Remediation

For the Enterprise

Per Department

1. Discover

- Content and Access Controls Scan
- Data Map
- Application Inventory
- Stakeholder and Authority Alignment

2. Plan

- Policy and Procedures Update
- Data Disposition Playbook
- Remediation Workplan

3. Remediate

- Detailed Content and Access Controls Scan
- Data Disposition
- Security and Access Remediation
- Data Migration

Technology Decisions

Program Governance

Change Management

Methodology for Content Remediation

For the Enterprise

Per Department

1. Discover

- Content and Access Controls Scan
- Data Map
- Application Inventory
- Stakeholder and Authority Alignment

2. Plan

- Policy and Procedures Update
- Data Disposition Playbook
- Remediation Workplan

3. Remediate

- Detailed Content and Access Controls Scan
- Data Disposition
- Security and Access Remediation
- Data Migration

Technology Decisions

Program Governance

Change Management

Example of Initial Scan Analysis

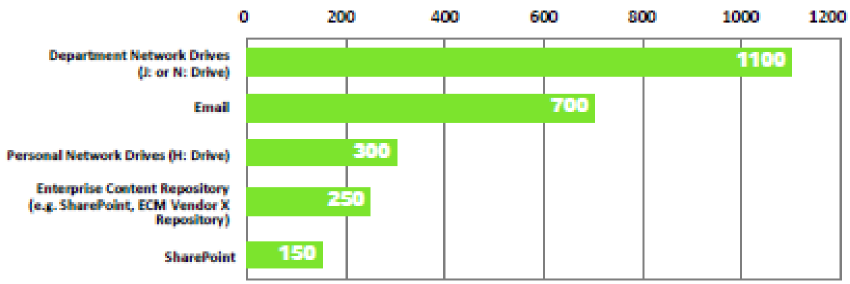
DETAIL: UNSTRUCTURED DATA

The unstructured content analytics and classification portion of the assessment provided accuracy rates that ranged from 57% to 96.%, depending on the data set. Most important was the collection of exemplar samples to use in the training process. For those departments or content samples where few samples or poor samples were provided, the results tended to be lower percentage in accuracy. Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx. The best opportunities exist within xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.

Where Are Documents Stored?

Doculabs worked with IT and Records Management to estimate the total amount of unstructured content stored in various systems at SampleCo. The results indicate xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.

Estimated Storage Consumption (U.S. Only, in Terabytes)



Assessment Scope

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Department/Function	Gigabytes	Classification Categories
Sales – Region 6	10	12
Finance	12	16
Customer Service	8	6
Application Support (IT)	15	9

Analytics Results

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

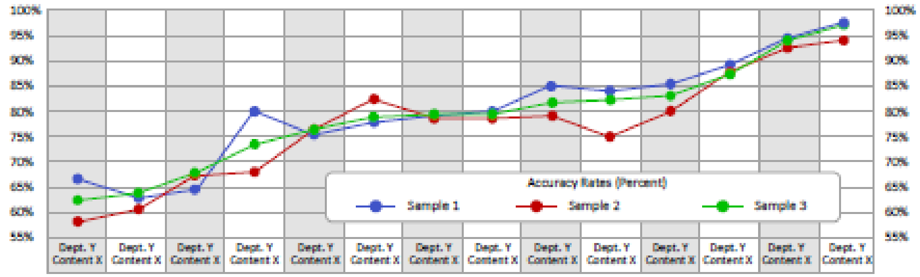
Preservation	Findings
Unnecessary File Types (Executables, Non-business Pictures, Movies, etc.)	13-15%
Duplicates	15-20%
Near Duplicates	9-30%

Risk	Findings
Files with PII	10-16%
Files with Sample Keywords	3-5%

Operational	Findings
Files 10 years or older	7-11%
Files accessed within the last 18 months	25-35%

Classification Accuracy Rates

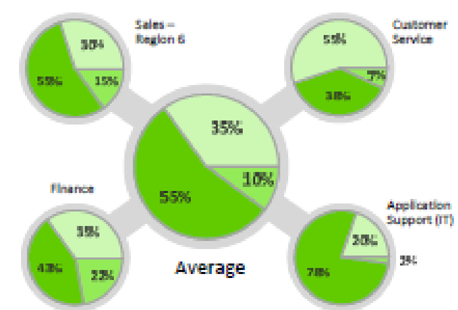
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx



Classification Results

In the graphic to the right, the portions of content classified as records, business reference, and of no value are shown for each function's content used in the assessment. Overall, 55% of the content has no value to the enterprise and roughly 10 were business records. xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx



Implications

The savings of $28 million are based on content growing 30% per year and SampleCo realizing a 20% per year reduction in the cost of storage due to negotiations with supplier X. In 2016, the 1,100 TB of content from 2012 will reach its disposition date.

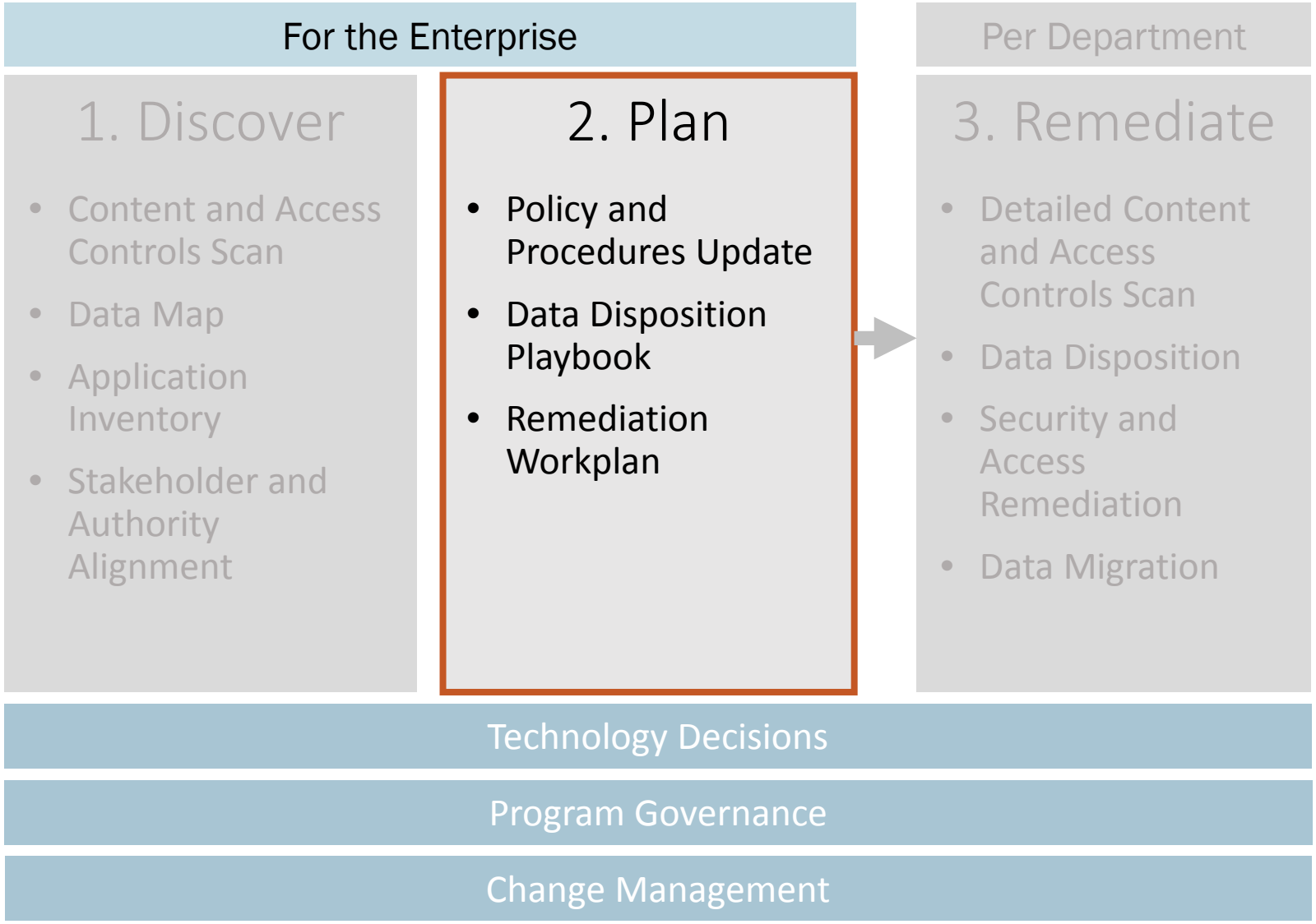
@$5,000,000 per PB	2012	2013	2014	2015	*2016	Total
Current Storage (PB)	2.50	3.25	4.23	5.49	7.14	
Current Cost ($Million)	$12.50	$13.00	$16.90	$21.97	$28.56	$92.93
Expected Storage (PB)	2.00	2.52	3.18	4.00	3.94	
Expected Cost ($Million)	$10.00	$10.08	$12.70	$16.00	$15.76	$64.55
Total Savings ($Million)	$2.50	$2.92	$4.20	$5.97	$12.80	$28.38

*In 2016, the 1,100 TB or 44% of content from the 2012 historical content assessment can be disposed

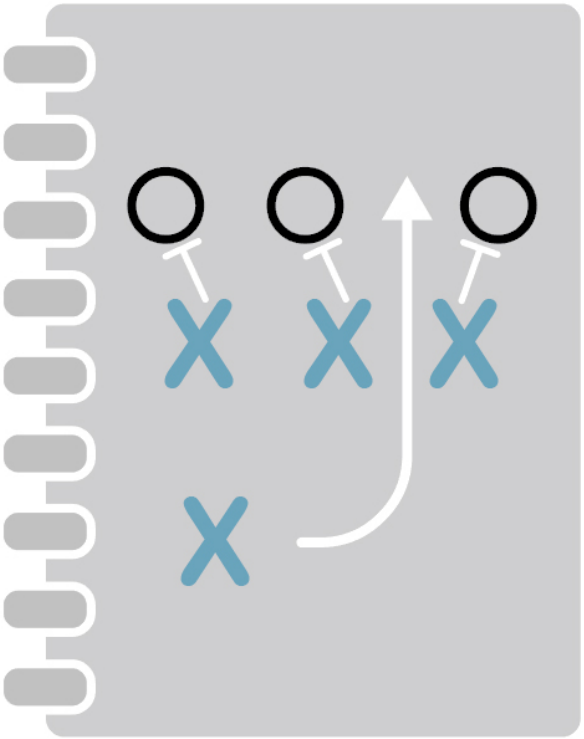
InfoSec and Records Management Alignment

- InfoSec is **beginning** to be concerned with information management
- While InfoSec has budget and organizational support, it (often) **lacks information management expertise**
- Records and information management can help InfoSec close this knowledge gap and be **successful**

Methodology for Content Remediation

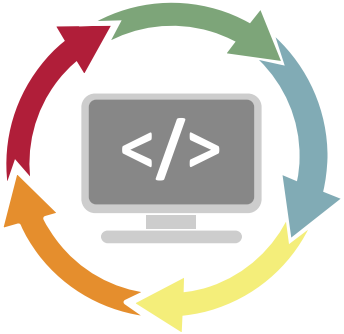


What is a Data Disposition Playbook?



- A way to document how business data was handled during an IT project so that an organization can explain how it did so at a later time

What Does a Playbook Look Like?



A modification of existing software development life cycle (SDLC) documentation to include more detailed, focused information management, records management, and legal information

OR

A single document that outlines the requirements, solution design, architectural decisions, test plan and results, and closeout steps to verify data was migrated/archived/deleted properly.

Why is it Needed?

- **IT** – to troubleshoot longer term issues, e.g., end users report problems with attachments 6 months after an email migration
- **IT** – to reuse methods from previous projects on current ones (don't reinvent the wheel)
- **Legal** – to determine how an organization handled data during an IT project to help them assess a current lawsuit or regulatory action and respond appropriately
- **Records and Information Management** – to make good information management part of the DNA of IT projects

How Can an Organization Use It?

- **Legal** – ten years on, if a lawsuit hits, legal can determine what exactly was done with data and be better prepared for early case assessment, meet and confer, depositions, etc.
- **IT** – lessons learned/troubleshooting, reuse of methods and procedures for on going projects
- **Information Management** – consistent participation in the SDLC process to ensure IM requirements and capabilities are delivered to manage data effectively
- **Records Management** – consistent participation in the SDLC process to ensure RIM requirements and capabilities are delivered to manage data effectively
- **Business** – gain understanding of how data is being handled in IT projects and applications

An Orphaned Data Policy

- If no data owner can be identified after a reasonable effort, information security (or information management) becomes the default owner and has the **responsibility and authority** to follow data preservation and disposition protocols
- Analogy: If a car was left in the corporate parking lot after work hours…
  - at 6 pm, building security looks up the license plate #, and if registered to an employee, an attempt is made to contact (ie “please remove your vehicle or we will have it towed”)
  - At 8 pm, if an employee does not contact security or if no one can be found associated with the vehicle, it is towed
- The vehicle represents a **potential threat or liability** to the headquarters facility, so action must be taken

Methodology for Content Remediation

For the Enterprise

Per Department

1. Discover

- Content and Access Controls Scan
- Data Map
- Application Inventory
- Stakeholder and Authority Alignment

2. Plan

- Policy and Procedures Update
- Data Disposition Playbook
- Remediation Workplan

3. Remediate

- Detailed Content and Access Controls Scan
- Data Disposition
- Security and Access Remediation
- Data Migration

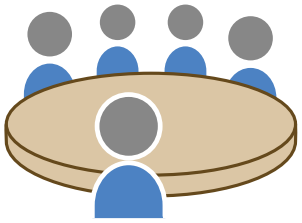
Technology Decisions

Program Governance

Change Management

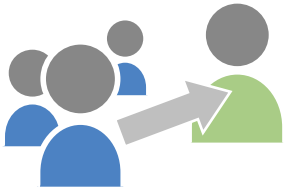
Data Cleanup and Disposition (Orphaned Data)

1



Conduct initial department content scan, use results to meet with department content owners to provide overview of scan results and agree on cleanup and disposition approach for orphaned data

2



Assign data ownership for information to be moved based on the clean up rules in initial department scan

3

Scan repositories based on established rules, move orphaned data to destination or temporary repository

4

Identify records in orphaned data set and flag with retention codes, migrate to records management system if appropriate

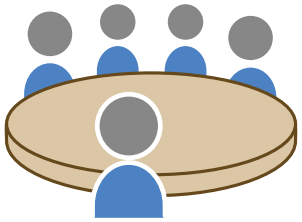
5

Identify legal holds on orphaned data set, ensure they are flagged or migrated to legal hold repository

6

Move remaining identified orphaned data into temporary archive or dispose per rules

Data Access Clean-up



1

Conduct initial department content scan, use results with department content owners to provide overview of initial scan results and agree on data access clean up approach

2 Identify folders in the source repository that have global access

3 Resolve broken inheritance and unresolved security identifiers

4 Remove global groups (for both stale data and active data)

5 Model correct access and permissions

6 Assign correct data owner

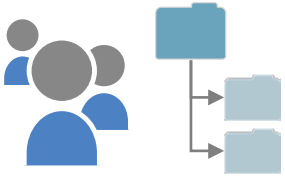
Steps 2 – 6 are repeated for each of the folders that requires changes in their access and controls.

Data Migration

1

Conduct clean up of current repository by identifying records, legal hold, ROT, etc.

2



Work with business, content, or record owners to develop proposed folder and metadata structure for SharePoint

3

Work with business, content, or record owners map to files and folders from source to target repository

4

Obtain business approval on source to target mapping, folder structure, and metadata

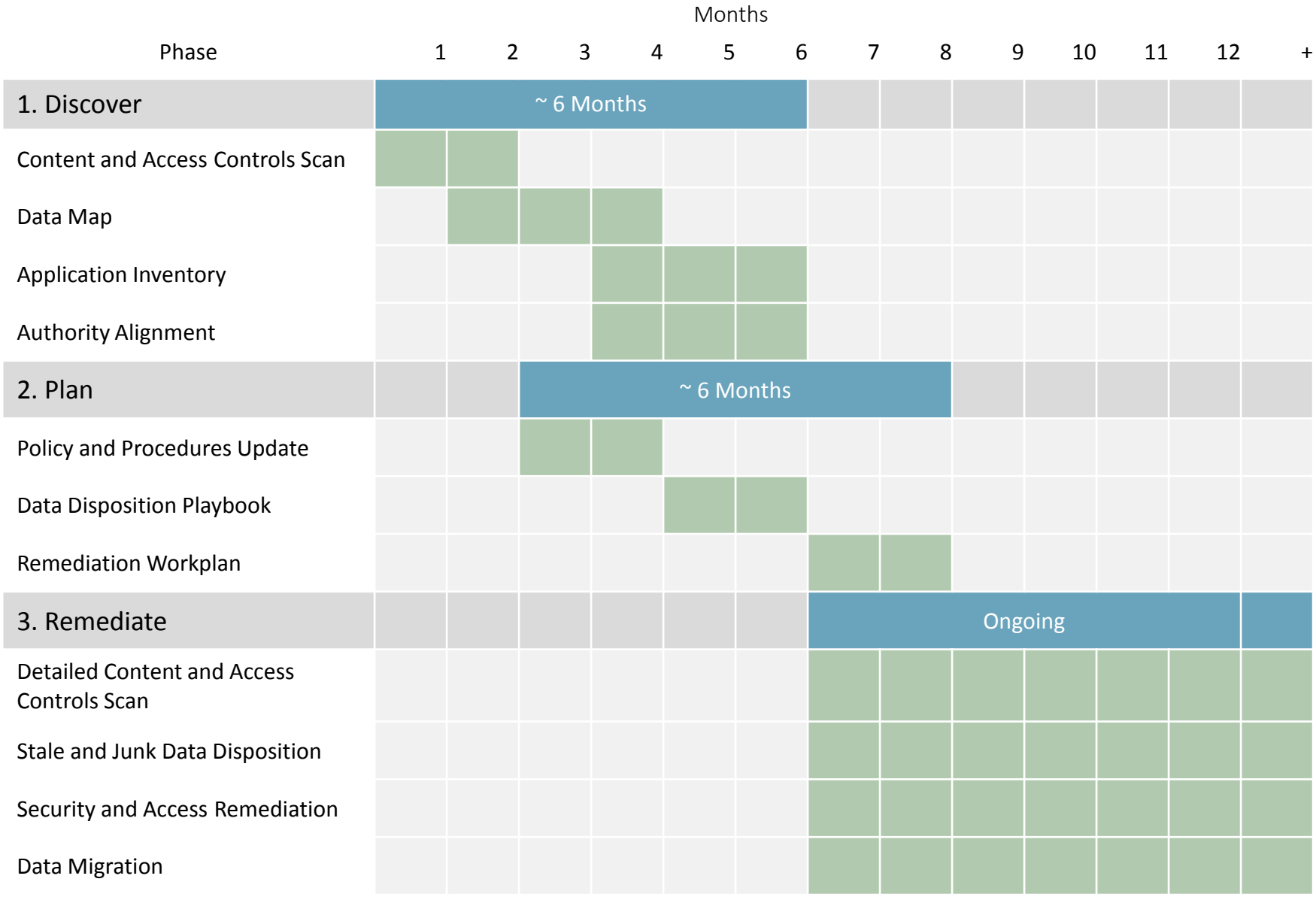
5

Using scripts or automated tools, migrate content from source to target repository (including source and target metadata)

6

Test the accuracy of content migration to target repository including source and target metadata assignment

Typical Timeline for Content Remediation



Change Management

- Change management is typically the **most overlooked** aspect of information management.
- Think about your organization's SharePoint rollout. Consider the level of change management support that was provided, and it's not difficult to see why the software may have provided **less value than anticipated**.
- You can't change the way everyone works with their day-to-day business documents without **over-communication** and **ample training**.

Agenda

- What's the problem?
- How do I solve it?
- What's in it for me?
- Bring it all together
- Q&A

Content Cleanup

The results of your repository scan are likely to be something like the following, which we've observed at dozens of clients over the last 10 years:

~30% to 70%

junk" content, which can be removed immediately

~20% to 40%

*stale content which can be archived or purged, depending on your approach

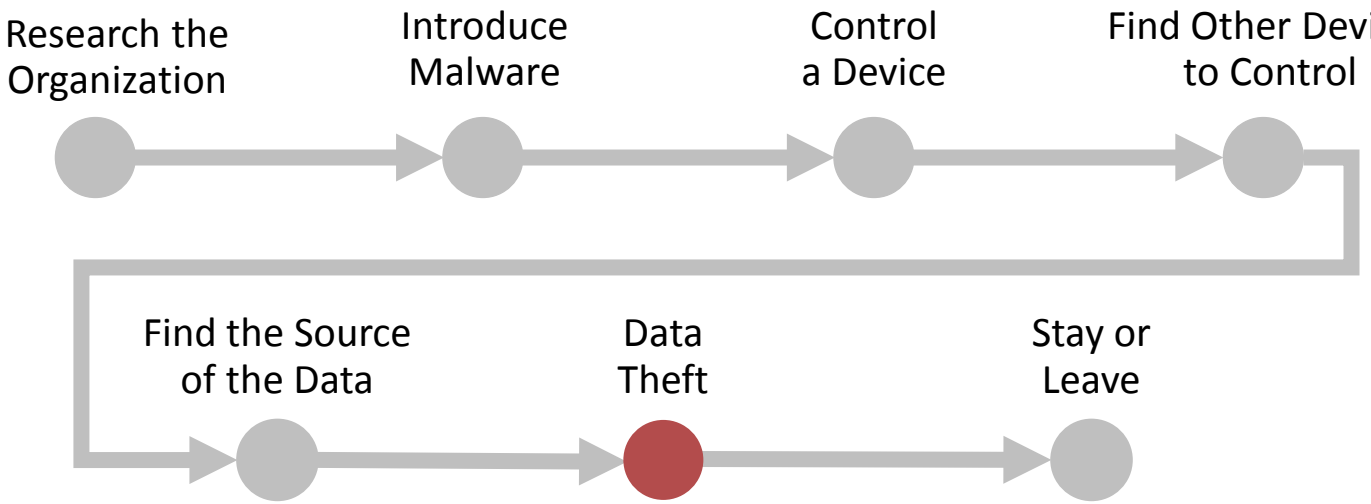
*(defined as older than 3 years, based on date last accessed),

~1 TB to 10TB

of stale sensitive content, which can be quarantined immediately with no operational impact

The Kill Chain: Reduce Data Theft

- Information security requires defending against what is often the weakest link in the cyberattack kill chain at organizations: **data theft**



Historically, data theft has been the weakest link in the Kill Chain, and Chief Information Security Officers (CISOs) are now turning to address it.

How does it help me?

Access

- "Just right" access to sensitive data leads to improved compliance with regulations such as HIPAA, HITECH, PCI, etc.
- Improve findability of content for employees (and employee efficiency), as they no longer need to sift through all the content they don't need.

Cost

- By reducing unstructured data an average organization with 100 TB of unstructured data, a 20% annual growth rate, and ~ 3% sensitive data can see cost avoidance of more than $14 million over 5 years.*
- Reduce the cost of outside counsel review of irrelevant data during eDiscovery. The cost of outside counsel review can be as much as $5,000-$30,000 per GB of data.**

Effort

- Legal and eDiscovery efforts are reduced during litigation events when content past its regulatory or business use life is disposed of.
- Information Security can more easily protect valuable/high-risk content.
- Reduced manual effort for managing content properly in managed repositories.

Risk

- By removing orphaned data you are likely to reduce the amount of fines in the event of a data breach.
- Well-managed and well-defined remediation initiatives show a "good-faith" effort with regulators and the public in the case of a breach.
- Increased effectiveness of audit and compliance functions for security and access.

* - Doculabs' *Information Security Calculator*, 2017

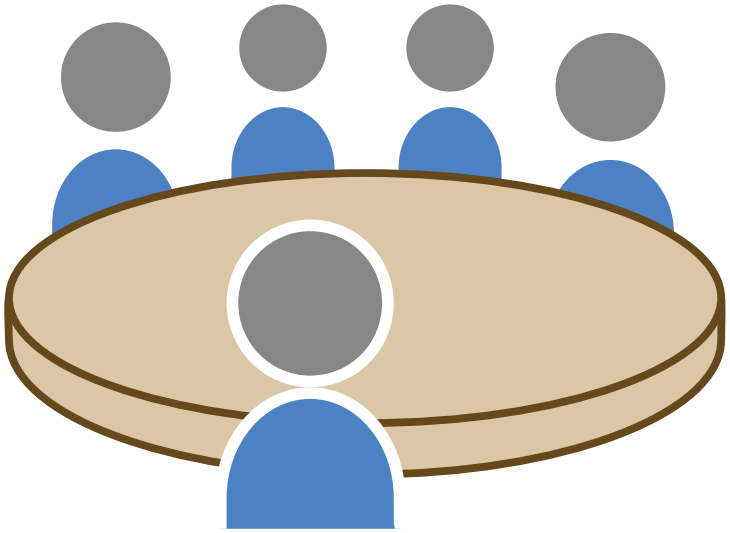
** - http://www.insidecounsel.com/2012/05/23/e-discovery-costs-pay-now-or-pay-later

Agenda

- What's the problem?
- How do I solve it?
- What's in it for me?
- Bring it all together
- Q&A

So Now What?

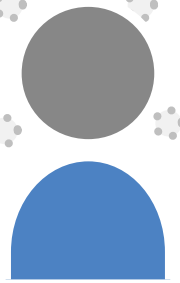
- Raise **awareness** in InfoSec about the importance of information management
- Articulate the **quick win efforts** InfoSec and Information Management can take:
  - Reduce junk and stale data
  - Identify sensitive data and take preliminary steps to protect it
- Reduce your **risk surface** and show progress to the C-level, the board, courts, and regulators



Ask Yourself

Do I know what data lives in what systems, who owns it, who has access to it, and who is accessing it?

Do I have agreement from key stakeholders on how to manage sensitive data, junk data, and stale data to reduce risk and increase value?



Do I have the policy and compliance infrastructure in place to allow me to manage data to reduce risk and increase value?

Do I have the technology in place to allow me to manage data to reduce risk and increase value in an efficient and sustainable way?

Questions?

Thank You

Matt McClelland

919-623-3663

mmcclelland@doculabs.com

Doculabs' Differentiators

- 25+ years in information management, records management, and e-discovery
- Broad customer base, with strong concentration in financial services
- Effectively bridges gaps between stakeholders (Legal, InfoSec, Records Management, IT, and the business), which is key to gaining consensus and driving decisions to move forward
- Deep experience with tools and techniques needed to execute sensitive data remediation, security and access remediation, and content migration

Want More Information?

- More information available on our website including:
  - White Papers
  - Blog
  - Information Security Calculator
  - Information on our services
- www.doculabs.com