



How to Clean Your Shared Drives & Build Effective Content Classification

Charmaine Brooks, CRM Partner

James Just, CDIA+ Partner

Question



Do you have a file plan,
taxonomy or classification
structure?

1. Yes
2. No
3. Don't know

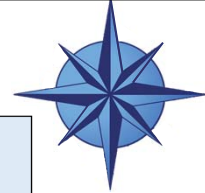
Semantic Schemes

Definitions



imerge

Semantic Schemes



Synonym Rings	Controlled Vocabulary	Taxonomy	Classification Scheme	Thesaurus	Ontology
Sets of words or phrases that can be used interchangeably for search.	Lists of preferred terms.	Systems for identifying, naming, and arranging things according to a set of rules.	An arrangement of knowledge, a complete ordered listing of all the items in a collection.	A tool that controls synonyms and identifies the semantic relationships among terms.	Faceted taxonomy with semantic relationships among terms and attributes with strict specification rules.
Relationships					
Equivalence		Hierarchy		Associative	



All are controlled vocabularies used in different ways in various systems.

Functional Classification

What is it?



imerge

Functional vs Organizational



Organizational Structures

grouped along organizational groups
(i.e. department)

not based on business functions or
activities, which often cross
organizational boundaries

Functional Structures

focuses on the high level purpose,
responsibility, or tasks for the
which the organization is
accountable

Functional Structure

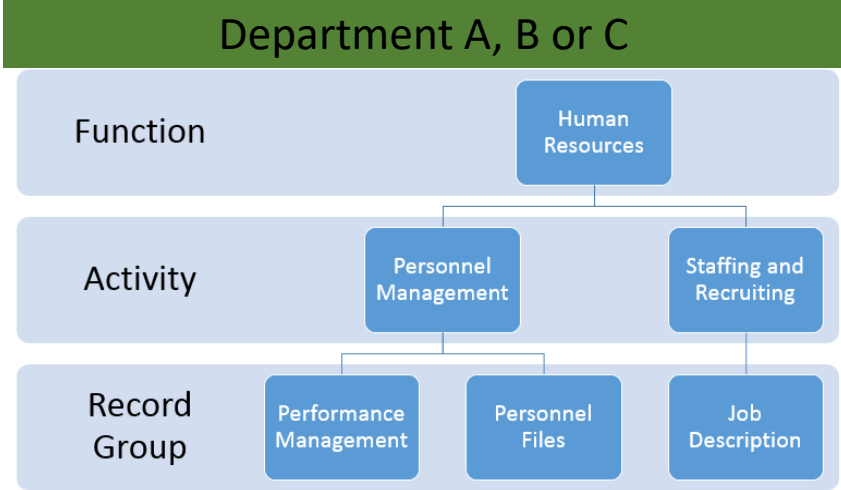


Functions Largest unit of business activity in an organization. Represent the major responsibilities that are *managed* by the organization to fulfill its goals. High-level aggregates of the organization's activities.

Activities/Sub-Activities are the next levels of the structure, taking the form of actions or transactions.

Record / Record Groups group similar records with the same retention requirements.

Functional Structure



Functions, Activities and Transactions



What your organization is mandated to provide.

Why the work is done to accomplish the mandate.

- What **tasks** support the mandate
- What **record or document groups** support the tasks

Functional Classification



Advantages

- More stable, don't change over time
- Records can be easily identified without affecting how they are classified
- Better search and retrieval within, and across, the organization
- Less reliance on cross referencing and scope notes
- Based on work processes

Purpose of Classification



Aggregation

- Establish a relationship between the Record and its context
- Link the Record to other related records

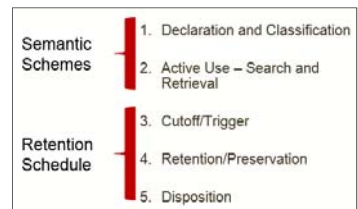
Search and Filtering

Discovery

Access and Security

Disposition Management

- Allow the Record to be managed effectively



Taxonomy vs Business Classification



Taxonomy the science of

- naming, describing and classifying things or concepts



Business classification used for

- findability and
- lifecycle management of content



Taxonomy vs Business Classification

	Taxonomy	Business Classification
Purpose	Categorization	Findability and Lifecycle Management
Designed and Managed By	Web Developers/ Scientists/ Librarians	Taxonomy Staff
Used by	Web Developers/ Scientists/ Librarians	Everyone and Computer Systems
Complexity	Detailed, Multiple Levels, Ontologies	Simple, seven to twelve top levels and no more than four levels deep
Key Characteristics	Exclusive, requires extensive knowledge	Usable, Intuitive, Common Vocabulary Easy to use with Minimal Training

Classifications Aren't Static

Giraffe aren't one Genus but four (and possibly 6)!



Scientific Classification

Kingdom	Animalia
Phylum	Chordata
Class	Mammalia
Order	Artiodactyla
Family	Giraffidae
Genus	Girraffa

Question



Do you have a records retention schedule?

1. Yes
2. No
3. Don't know

What is a Retention Schedule?



Record Series/Type

- Groups of similar records with the same retention requirements

Retention Period

- Minimum/maximum time period records need to be retained

Trigger date

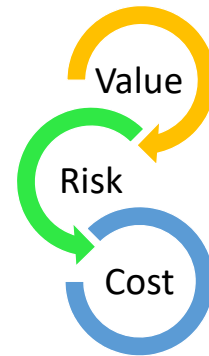
- Date that triggers retention and disposition

Trigger event

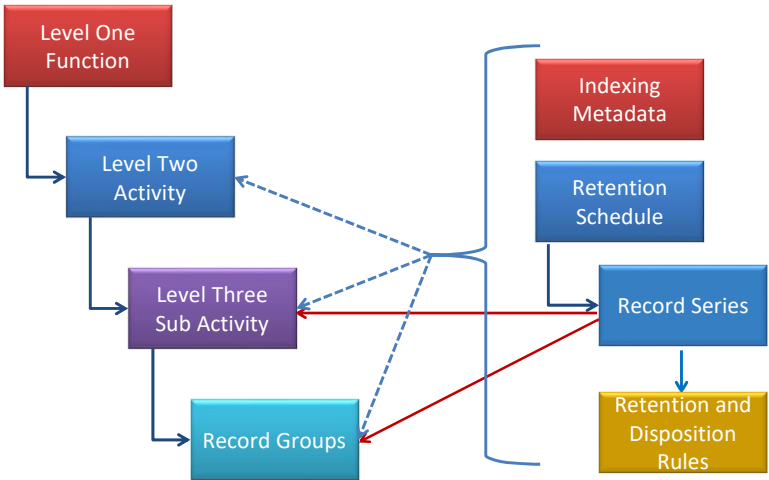
- If the retention is conditional on a specific event, what is that event (e.g. termination of employee, completion of contract)

Authority for retention / disposition

- What legal citation is the authority for the time period cited
- If there is no legal citation, what is the time based criteria



ISO 15489 Classification and Retention



Classification

Exercise



Classification Structure



Shared Drive Clean Up

Unstructured Records



imerge

21

Question

Have you tried to clean up a shared drive?



1. Yes
2. No
3. Don't know

Question



Was it easy and successful?

1. Yes
2. No
3. Don't know





**WILD
UNTAMED
UNCONTROLLED**

Question



What is the volume of your shared drives?

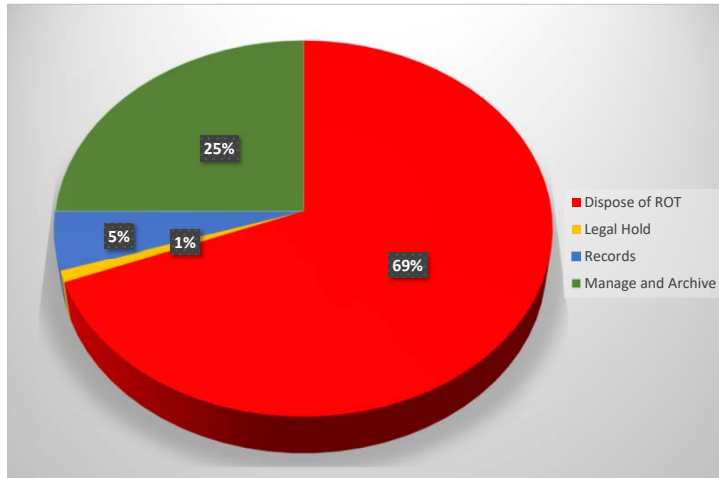
1. 1 TB to 10 TB
2. 10 to 50 TB
3. 50 TB+
4. Geez, I wish I knew



QTY of BUSINESS DOCUMENTS?

Not backups, programs, source code, etc.

Why Remediate Shared Drives?



Why Remediate Shared Drives?



Reduce Costs

- Never delete? Space & costs increase logarithmically

Efficient retrieval of information

- Eliminate clutter
- Structure content logically (or move to ECM/SharePoint)
- Appropriate tagging results in information findability

Encouraging information sharing

- Well organized shared drives (or use of ECM/SharePoint) simplifies security and sharing of information across the organization

Why Remediate Shared Drives?



Reducing redundancy and versions

- Improved information sharing reduces duplication
- Reduces the risk of using wrong version for business decisions

Apply Retention Disposition Schedules

- Mixed-lifecycle content in folder structures
- Applied consistently will likely reduce storage requirements and slow the growth curve

Remediation Analysis Activities

imerge

General Content Grouping



ROT

- Redundant, Outdate, Trivial/Trash

High Value

Low Value

Unknown

Overarching Consideration



Capstone status of content creator

- Mayor v. HR Director v. Admin Asst

Links

- Document links will break when content is moved/deleted

Special Consideration

- Holds (litigation, audit, investigation)
- Archival value (historical value)

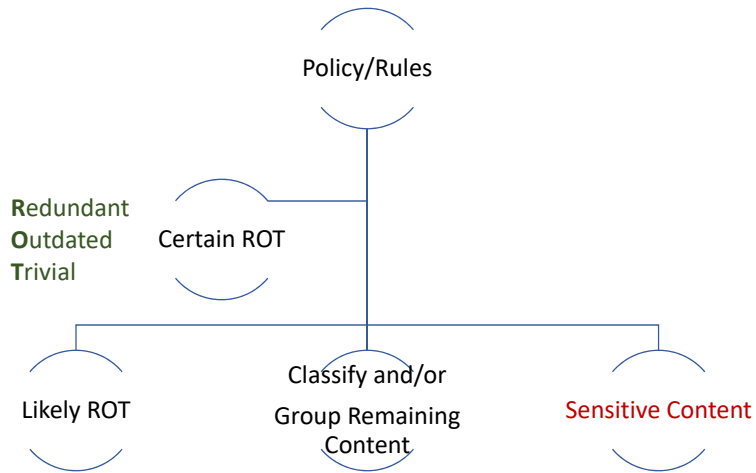
Overarching Consideration



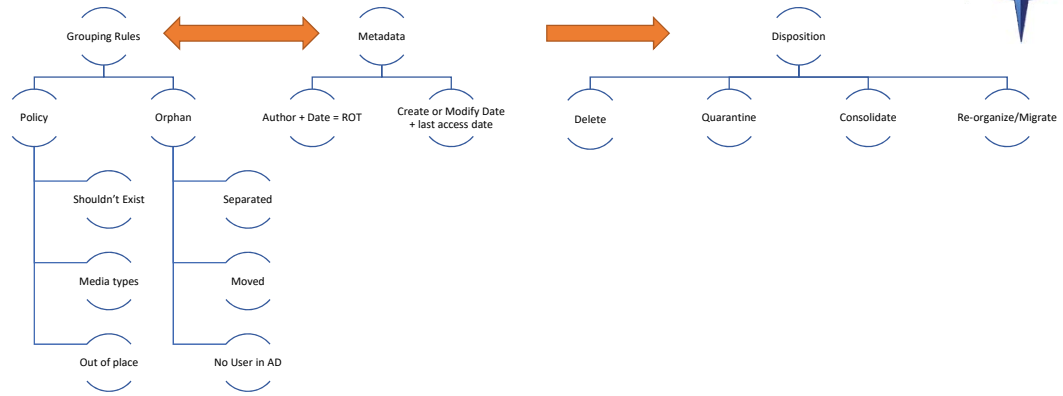
Risk

- Risk of keeping longer than retention rules
- Cost of keeping longer than retention rules
- Risk/cost of erroneous decision
- Spoliation

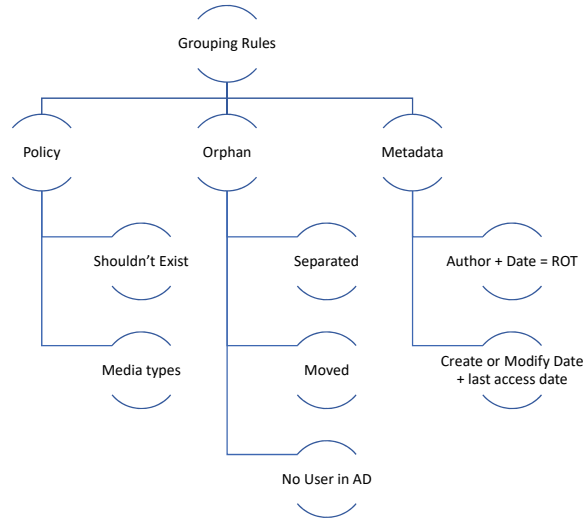
Basic Analysis Workflow



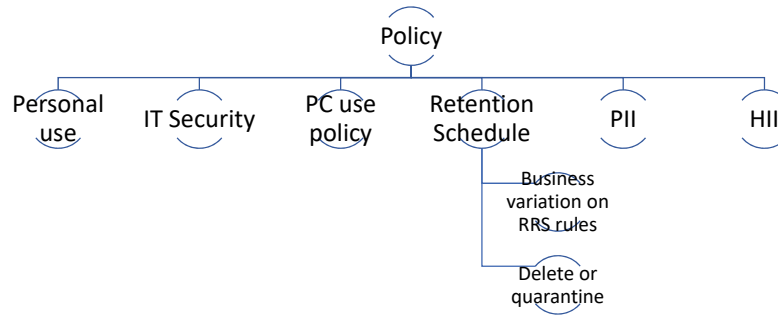
General Decision Tree



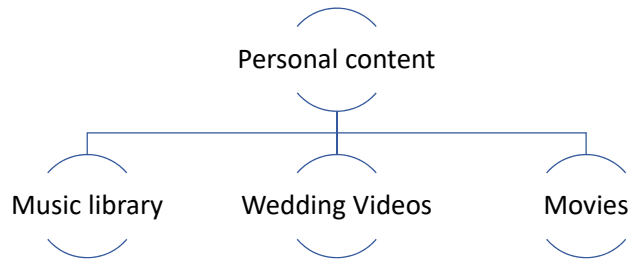
Grouping Content for Action



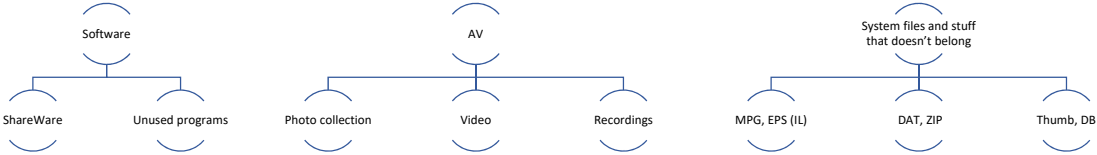
Policy Drives Decisions



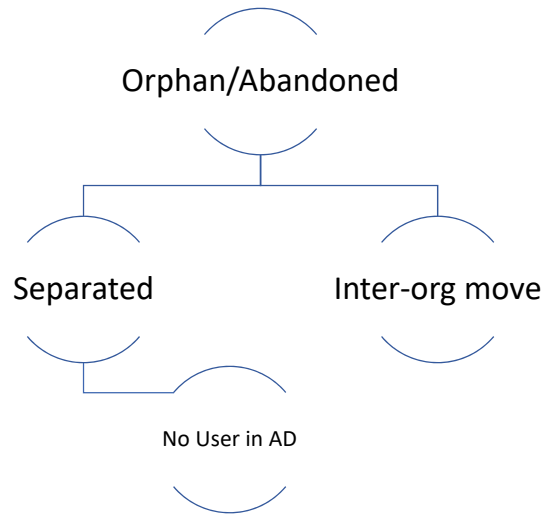
Policy: Shouldn't Exist



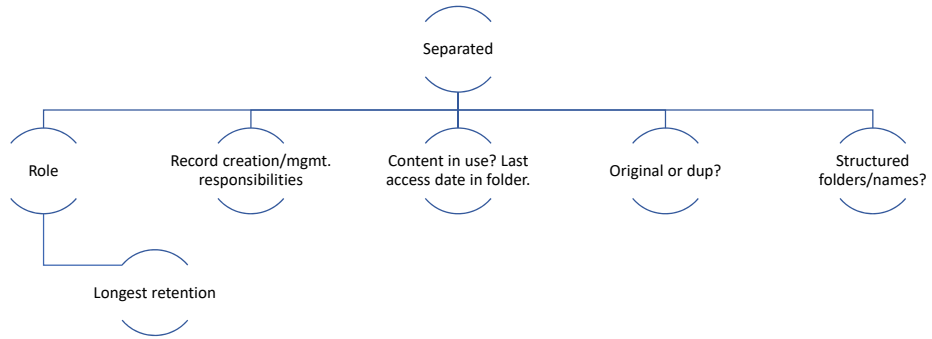
Policy: Media types



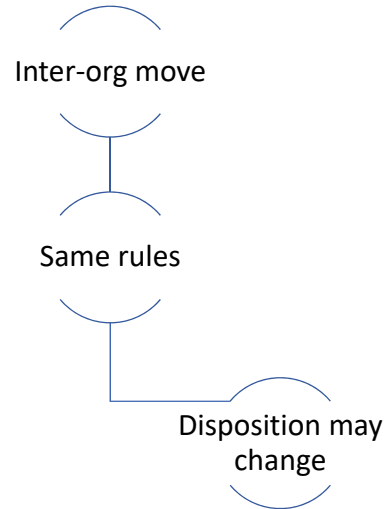
Policy: Orphan/Abandoned



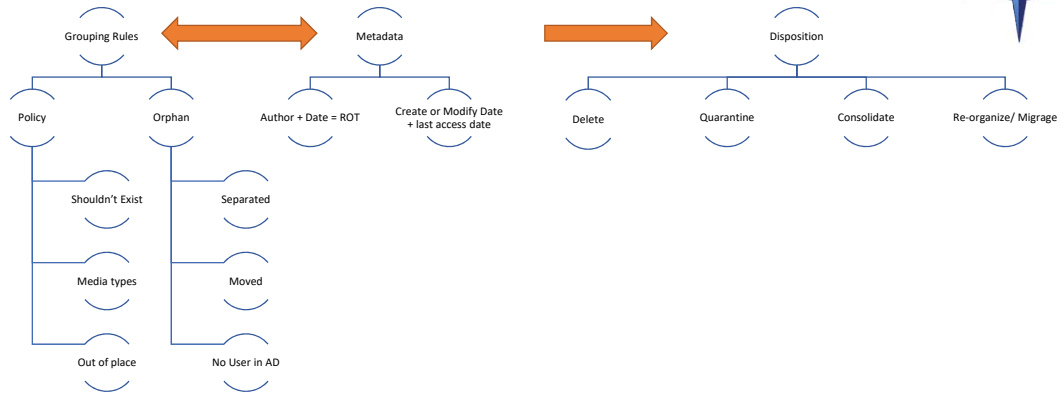
Policy: Orphan/Abandoned



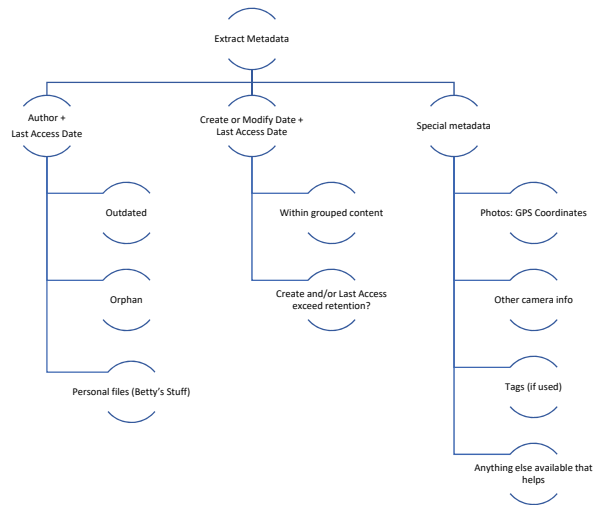
Policy: Orphan/Abandoned

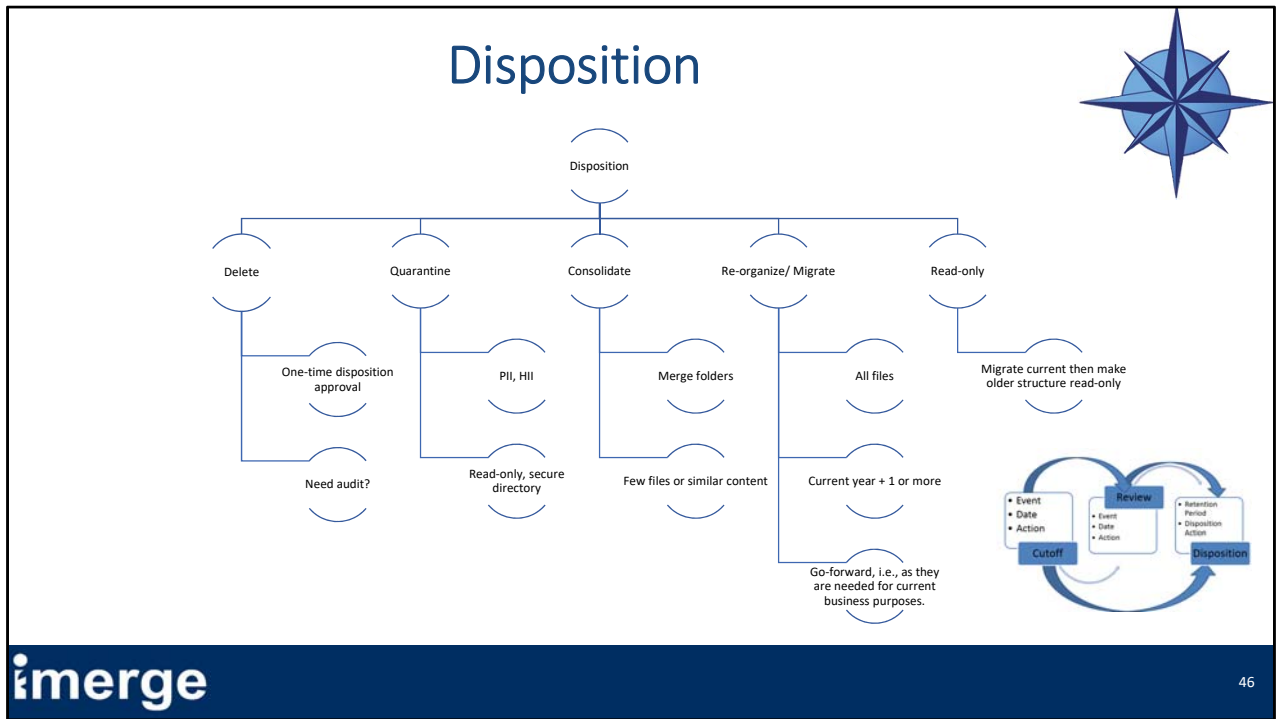


Decision Tree



Metadata Analysis





Rather than deleting or migrating documents (not ROT), especially more recent content, move it to a “to be deleted” folder. When migrating or deleting, links in documents will be broken. Use a link-finding software program before migrating content.



imerge

Examples



<input type="checkbox"/> Name	Date modified	Type
Admin	11/17/2015 10:05 ...	File folder
Betty's files	4/11/2005 4:29 PM	File folder
Expense reimbursements	4/11/2016 11:27 AM	File folder
Vendors	11/17/2015 10:06 ...	File folder

- Betty left the city in 2006
- Longest retention rule in her department is 7 years
- Last modify or access date is over 10 years ago
- Delete it unless Capstone status.

- Expenses organized by year
- Expense retention is 5 years
- Delete 2012 and earlier

▼ Finance

- Admin
- ▼ Expense reimbursements
 - 2010
 - 2011
 - 2012
 - 2013
 - 2014
 - 2015

Dealing with Duplicates



Duplication

- Name, size, date and time match - Positive duplicate
- Name, size and date match – Probable duplicate
- Name and size only match – Likely duplicate
- Name only match – Not duplicate
- Content Analysis systems use Hashing

Dealing with Duplicates



Define how duplicates will be handled

- Keep the newest file and place the rest in duplicate folders (positive and probable duplicates *in the same folder*)
- Understand why the duplication exists and determine if duplicates should be kept or discarded
- Non-duplicates with the same name - Append file names based upon current folder information if helpful or append with number

Example



File Path and Name	File Size	Last Access	Action
w:\Safety, Security and Environmental\Accident Info\02-02-11.doc	70,144	2/1/2011	
w:\Safety, Security and Environmental\Accident Info\02-16-11.doc	70,144	2/15/2011	
w:\Safety, Security and Environmental\Accident Info\03-02-11.doc	68,608	3/1/2011	
w:\Safety, Security and Environmental\Accident Info\03-16-11.doc	69,120	3/15/2011	
w:\Safety, Security and Environmental\Accident Info\03-30-11.doc	68,096	3/29/2011	
w:\Safety, Security and Environmental\Accident Info\Thumbs.db	5,632	1/27/2011	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Clean June 1 2009.doc	76,288	6/1/2009	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Legal.doc	61,440	5/14/2009	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Legal and mine.doc	62,464	5/14/2009	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Policy Revision 09.doc	129,024	4/30/2009	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\process- update 2007.doc	98,816	10/30/2007	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Appeals Determination Form.doc	22,016	5/4/2007	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\ 1-8-10 review and determinations.doc	29,184	1/8/2010	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\1-15-10 review and determinations.doc	35,328	1/18/2010	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\1-29-10 review and determinations.doc	29,696	1/29/2010	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\1-29-10 review and determinations.docx	16,551	2/1/2010	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\02-12-10 review and determinations (2).doc	33,280	2/15/2010	

Example



File Path and Name	File Size	Last Access	Action
w:\Safety, Security and Environmental\Accident Info\02-02-11.doc	70,144	2/1/2011	
w:\Safety, Security and Environmental\Accident Info\02-16-11.doc	70,144	2/15/2011	
w:\Safety, Security and Environmental\Accident Info\03-02-11.doc	68,608	3/1/2011	
w:\Safety, Security and Environmental\Accident Info\03-16-11.doc	69,120	3/15/2011	
w:\Safety, Security and Environmental\Accident Info\03-30-11.doc	68,096	3/29/2011	
w:\Safety, Security and Environmental\Accident Info\Thumbs.db	5,632	1/27/2011	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Clean June 1 2009.doc	76,288	6/1/2009	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Legal.doc	61,440	5/14/2009	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Legal and mine.doc	62,464	5/14/2009	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Policy Revision 09.doc	129,024	4/30/2009	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\process- update 2007.doc	98,816	10/30/2007	
w:\Safety, Security and Environmental\Accident Info\Accident Review & Appeal policy\Appeals Determination Form.doc	22,016	5/4/2007	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\ 1-8-10 review and determinations.doc	29,184	1/8/2010	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\1-15-10 review and determinations.doc	35,328	1/18/2010	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\1-29-10 review and determinations.doc	29,696	1/29/2010	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\1-29-10 review and determinations.docx	16,551	2/1/2010	
w:\Safety, Security and Environmental\Accident Info\Accident Reviews\2010\02-12-10 review and determinations (2).doc	33,280	2/15/2010	

Dealing Versions and Drafts



Understand Version/Draft Rules

- Requirement for keeping versions/drafts? (Need to know how the sausage was made.)
- Retention rule for version/drafts
- Business process for cleaning up versions/drafts when final approved
 - Using SharePoint workflow for the review process can automate cleanup once a “final” has been designated
 - Using simultaneous editing or linking to a single document simplifies management



Business Rules Rule

imerge

Case and Project files



Case and project files

- Longest retention on any document or
- Retention on the overall file as determined by rule
- Business requirement



imerge

Define Business Rules



Remediation Policies

- Duplicate handling
- Orphan data
- Quarantine or migrate
- Records that have met retention

Define Business Rules



Criteria

- Author/Title
- Create Date > xx/xx/xxxx
- Last Access Date > xx/xx/xxxx
- Could use content words or phrases

Get Policy Approved

City of Madison Records Retention & Disposition Schedule

Records Series No.	Record Series Title	Status	Format	Retention	Disposition	Description
GENERAL RECORDS SERIES - All Agencies						
GS000001	ACCOUNTS RECEIVABLE/ACCOUNTS PAYABLE	Active	Paper, Electronic	FIS+7 years	Destroy - Destroy Conf	Records related to accounts receivable and accounts payable. May include write-offs, purchase orders, payment history records, invoices, receipts, vouchers, cash receipts, deposit forms, receipts for parking tickets, taxes, or background checks. Receipts may contain personally identifiable information.
GS000015	DEPARTMENT ANNUAL REPORTS	Active	Paper, Electronic	Permanent		Summary reports documenting the program activities and accomplishments of City departments and divisions.
GS000018	GENERAL CORRESPONDENCE	Active	Paper, Electronic	CR+7 years	Transfer* to WHS or Destroy-Confidential	General correspondence generated or received by City staff and not directly related to a program. May contain personally identifiable information. <i>*Mayor's Office & Dept/Div Heads correspondence is transferred to WHS.</i>
GS000030	CONTRACTS	Active	Paper, Electronic	EVT+10 years EVT-End of contract term	Destroy-Confidential	Contracts executed by the City of Madison. Includes, but not limited to, memorandums of understanding; public works, purchasing, service, or development contracts; inter-governmental, advertising or cooperative agreements. Attachments or exhibits to the contracts may contain trade secrets or personally identifiable information.
GS000038	HIRING AND RECRUITMENT	Active	Paper, Electronic	FIS+3 years	Destroy-Confidential	Hiring and recruitment documentation. May include requisitions, job postings, applications, resumes, test scores, interview notes, or correspondence.
GS000066	PROCUREMENT REQUEST FOR BID/PROPOSAL/INFORMATION	Active	Paper, Electronic	FIS+7 years	Destroy-Confidential	Documents related to the procurement of goods and services which may require a request for information, bid, or proposal. May include committee activities, official request, record of award, amendments, protests, disputes, complaints, correspondence and related records.

Define Business Rules Agency All Retention Schedule

Record Series	Retention	Business Rule
Accounts Payable	FIS + 7 years	
Contracts	End of Contract + 10 years	
Department Annual Reports	Permanent	
General Correspondence	CR + 7 years	
Hiring and Recruitment	FIS + 3 years	
Procurement RFB/P/I	FIS + 7 years	

Business Rules 2018

Record Series	Retention	Business Rule
Accounts Payable	FIS + 7 years	Invoices, PO's prior to 2010
Contracts	End of Contract + 10 years	Closed contracts prior to 2007
Department Annual Reports	Permanent	Do not destroy – move to Archive
General Correspondence	CR + 7 years	Correspondence prior to June 2011
Hiring and Recruitment	FIS + 3 years	Job postings, applications prior to 2014
Procurement RFB/P/I	FIS + 7 years	Bids prior to 2010

On-going Monitoring and Maintenance



- Changes to the shared folder structure require the agreement of the Records Liaison and Records Manager or his/her delegate and should be documented as amendments to the Classification Structure. Only the person authorized to do so may add, delete, or alter classifications.
- The Records Liaison should monitor activity on the shared folder structure to ensure that users are following the documented business rules and naming conventions.
- Monitoring much easier with content analysis tools

Migration



imerge

Migration: Prerequisites



Governance

- Authorities, roles and responsibilities for the capture, control, disposal, access and storage of records. Governance should include defining the responsibilities between management, stakeholders, end users, and systems administrators.

Migration: Prerequisites



Governance: RACI

- Responsible
- Accountable
- Consulted
- Informed

Migration: Prerequisites



Governance: RACI

Role	Record Creators	Operational Record Owners	Record Custodians	Business Users
Organizational Role	Manger Specialists Administrators Analysts Assistants	Specialists Administrators	Administrators Assistants	Analysts Assistants
Name	A	C	C	I
RACI				
Responsibility				

Migration: Prerequisites



Create classification scheme/taxonomy

- ISO15489 Functional Classification
- Departmental file plans
- One retention rule for a folder structure
- Differentiate drafts and final versions
- Security – promote sharing, eliminate duplicates
- Standardize metadata (folder and file names)

Taxonomy or Defined Classification Structure

- Finance and Accounting
 - Accounting Management
 - Accounts Payable
 - Check Requests
 - Expense Reports [Year] [Employee]
 - Invoices Year] [Vendor]
 - Credit Card Statements [Year] [Employee]
 - Accounts Receivable
 - Credit and Collections
 - Statements [Year] [Customer]
 - Budget Management
 - Approved Year]
 - Development and Preparation [Year]
 - Procurement Administration
 - Bids and Proposals [Prepmt] [Close/Completion Date]
 - Catalogs and Pricing
 - Purchase Orders and Requisitions [Number] [Prepmt] [Close/Completion Date]
 - Vendors, Contractors and Suppliers [Name] [Close/Completion Date]
 - Governance and Compliance
 - Audit Management
 - External
 - Financial [Year]
 - Internal
 - Business and Strategic Planning
 - Departmental
 - Organizational
 - Committee and Meeting Management
 - Meetings and Briefings --
 - External [Type] Meetings
 - Internal Departmental
 - Policy, Procedure and Standards Management
 - Guidelines and Manuals [Title] [Superseded Date]
 - Policies [Policy Number] [Superseded Date]
 - Procedures [Title] [Superseded Date]
 - Signing Authorizations [Title] [Superseded Date]
 - Standards and Regulations [Title] [Superseded Date]
- Human Resources
 - Employee Development
 - Performance Appraisals and Objectives [Year] [Employee ID/Name]
 - Performance Goals and Review [Year] [Employee ID/Name]
 - Personnel Administration
 - Personnel Files [Employee ID/Name] [Term Date]
 - Time and Attendance
 - Work Schedules and Assignments
 - Staffing and Recruiting Administration
 - Job Descriptions [Position Name] [Superseded Date]
 - Recruitment and Selection
 - Applications - Not Hired [Year]
 - Legal
 - Contract Management
 - Goods and Services [Contract Number] [Expiration Date]
 - Employment [Employee ID/Name] [Term Date]
 - Records Governance
 - Classification Scheme [Superseded Date]
 - Disposition Destruction Approvals and Certificates
 - Inventories and Surveys Offsite
 - Legal Hold [Case Name] [End Date]
 - Public Records Requests [Request Number] [Year]
 - Retention Schedule [Superseded Date]
 - Risk, Safety and Security
 - Insurance
 - Policies [Expiration Date]
 - Safety and Security Management
 - Emergency Management
 - Continuity of Operations
 - Plans, Safety

Folder Structure - Folder/File Naming



Business Area or Department A, B, C

- Function
 - Activity
 - Sub-folders
 - Transaction/Title
 - Case Files (provides metadata such as project ID)

Folder Structure - Folder/File Naming



Name documents and files consistently and logically

- Consistency is the key (i.e. project numbers, vendor names, etc.)
- Date formats (YYYYMMDD) is best practice

Names should be self-explanatory and meaningful

- Do not use “general”, “misc” or “personal names” – these are SHARED drives

Don't repeat previous folder level names

Folder Structure - Folder/File Naming



Organizing information for lifecycle management

- By year for time-based retention
- By trigger event for event based retention (year inactive, completed, expired, etc.)

Meaningful Naming Conventions

- Records should not be stored by file format
- e.g. do not store spreadsheets in a folder called "Spreadsheets"

Manual Migration



New taxonomy – Messy to Clean

- Migrate old folder structure to new, “clean”
 - Or, migrate to ECM/SharePoint
 - Or, build new structure and make old “read only”
 - Or, combination of these

Migrate to SP



Metadata available in folder and file names

- Example: Contracts
 - Folder Name: 12345 Smith Cleaning Contract 2014-01-01
 - File Name: 12345 Contract 2014-01-01
 - File Name: 12345 Contract Attachment A 2014-01-01
 - File Name: 12345 Addendum 1 2014-04-15

ECM/SharePoint Migration rules

- Extract Contract number 12345 from Folder Name
 - Validate against contract tracking Access database
 - Pull party names from Access
- Extract document type from File Name
- Extract document date from File Name

Content Analysis Software for Shared Drive Remediation

imerge

Question



Does your organization have tools for file analysis and clean up?

1. Yes
2. No
3. Don't know

Summary



- Planning and project management is critical
- Ensure resources are available and trained
 - Staff and working teams
 - Tools and technology
- Develop a process and test, test, test
- Provide support throughout the process and on an ongoing basis
- It will take time!!! Be realistic!

How to Clean Your Shared Drives & Build Effective Content Classification



Charmaine Brooks, CRM
James Just, CDIA

IMERGE Consulting, Inc.



James Just, Partner

James.Just@imergeconsult.com

(608) 239-8282

Charmaine Brooks, CRM Partner

Charmaine.Brooks@imergeconsult.com

(208) 867-8337