



e-Nterprise  
Advisors

# Classification and Metadata

**Priscilla Emery – President  
e-Nterprise Advisors**

# Agenda

- **Why Classification Schemes are important.**
- **Differences between functional and hierarchical classification schemes.**
- **How a classification scheme is used within a RIM program and beyond.**
- **The role controlled vocabularies and metadata have in supporting classification schemes and models.**

# Taxonomy: An Important First Step

- Taxonomy is the science of classification according to a pre-determined system, with the resulting catalog used to provide a conceptual framework for discussion, analysis, or information retrieval
  - Scientific classification of items into specially named groups based on shared relationships
  - A hierarchical list of terms or content categories
  - A consistent method of organizing stored knowledge so you can quickly find the information you need
- May need to rationalize different internal classification schemes.

# Classification Scheme

- A way to describe an organization's information landscape by identifying key business functions, processes and information types
  - **Function** - business activity conducted in support of an organization's mission, i.e., what needs to be accomplished
  - **Process** - activity conducted in support of a function, i.e., how it gets accomplished
  - **Information Type** - group of related records created, managed, maintained, stored or disposed of as part of a function or process, i.e., the results (or inputs and outputs) of what's accomplished

# The Importance of Classification

- **Classification Provides Context of Records**
  - Search engine indexing does not provide context
- Classification “up front” in document management systems can simplify eventual retention and disposition of records in the future
- Can make it easier to scope out and find records during an audit or Discovery activity

# Importance of Classification Schemes

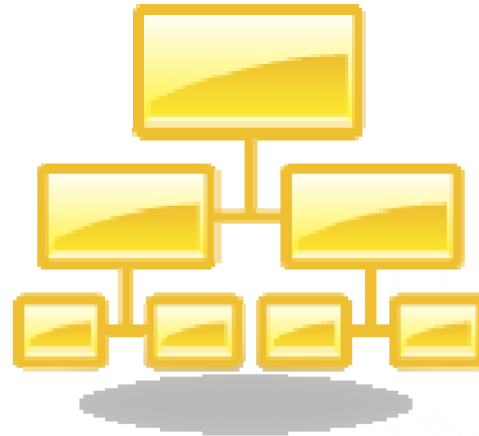
- Makes it easier to apply and implement retention
  - Streamlines record categories
  - Reduces time required to maintain the retention schedule
  - Results in a more user-friendly retention schedule

# Importance of Classification Schemes

- Standardizes how everyone organizes information in all repositories
- Basis for consistent folder structure
- Facilitates searching
  - Creates a standard set of metadata tags
  - Reduces time to locate required information

# Classification is Not a One Size Fits All Proposition

- [Classification Video](#)



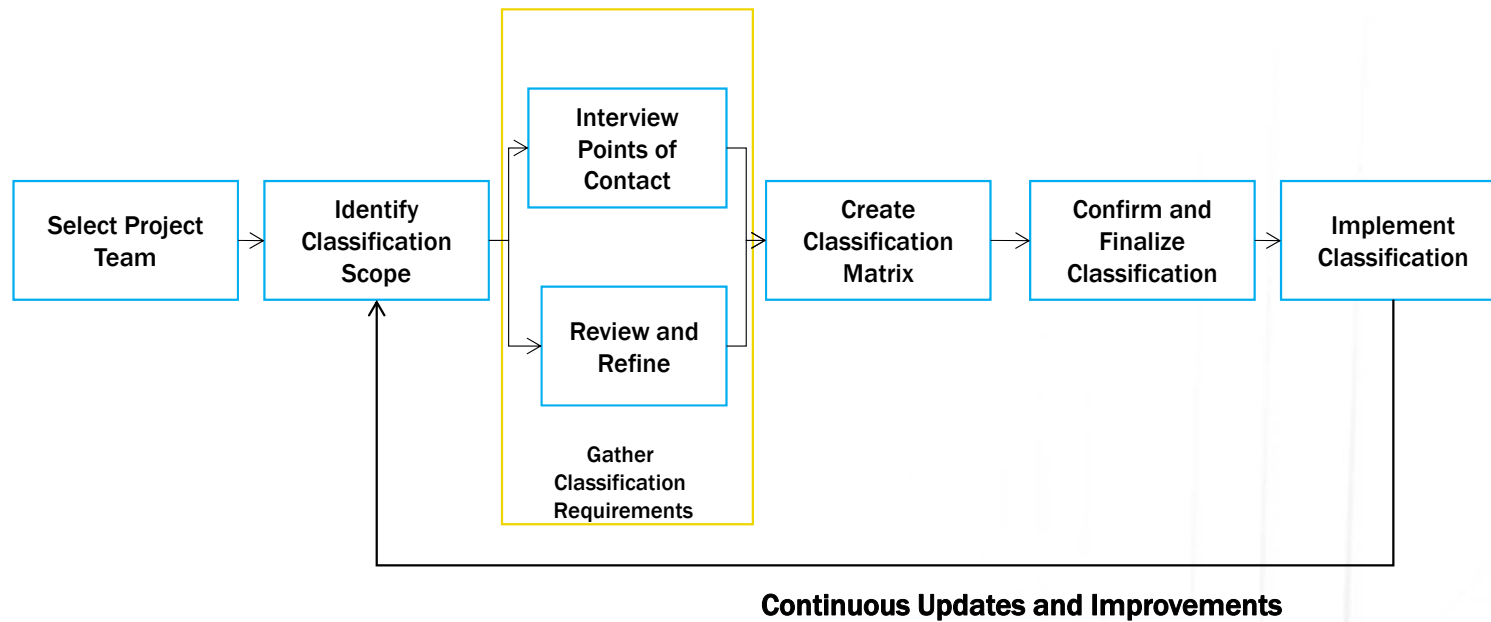


# 9 Principles of Classification

1. Intuitive
2. Unambiguous
3. Hospitable (i.e. welcoming of change)
4. Consistent and Predictable
5. Relevant
6. Parsimonious
7. Meaningful
8. Durable
9. Balanced

Source: AIIM International

# Developing a Classification Scheme



# Classification Scheme Project

## ➤ Select classification team

- **Records Management** – knows the end user and what information is out there
- **Information Technology** – knows what an organization's technology can and cannot do
- **Legal and/or Compliance**– knows litigation, regulatory environment and impact on the organization

# Steps to Develop a Classification Scheme

- **Identify classification scope**
  - **Agree on purpose and scope and tie the classification project into other organizational strategic objectives**
  - **Make your business processes more efficient; identify interactions among people, process, technology and information**
  - **Identify your intellectual property; identify information assets and who can access**

# Steps to Develop a Classification Scheme

- **Identify classification requirements**
  - **Gather information on organization and related documents (e.g., retention schedules, file plans)**
  - **Identify key business Points of Contact (POCs) from all business units**
  - **Develop interview guide**

# Steps to Develop a Classification Scheme

- **Identify classification requirements**
  - **Conduct interviews**
    - **Send out e-mail introducing project and what you need from the POCs**
    - **Be sure to define unfamiliar terms – don't assume everyone speaks "records management"**
    - **Make decisions on follow up interviews – you can't interview everyone**

# Steps to Develop a Classification Scheme

- Identify classification requirements
  - Document interview results by developing draft classification schemes
    - Functions, processes and information types
    - Concerns or issues
    - Retention needs
  - Format can be any layout – keep it simple

# Steps to Develop a Classification Scheme

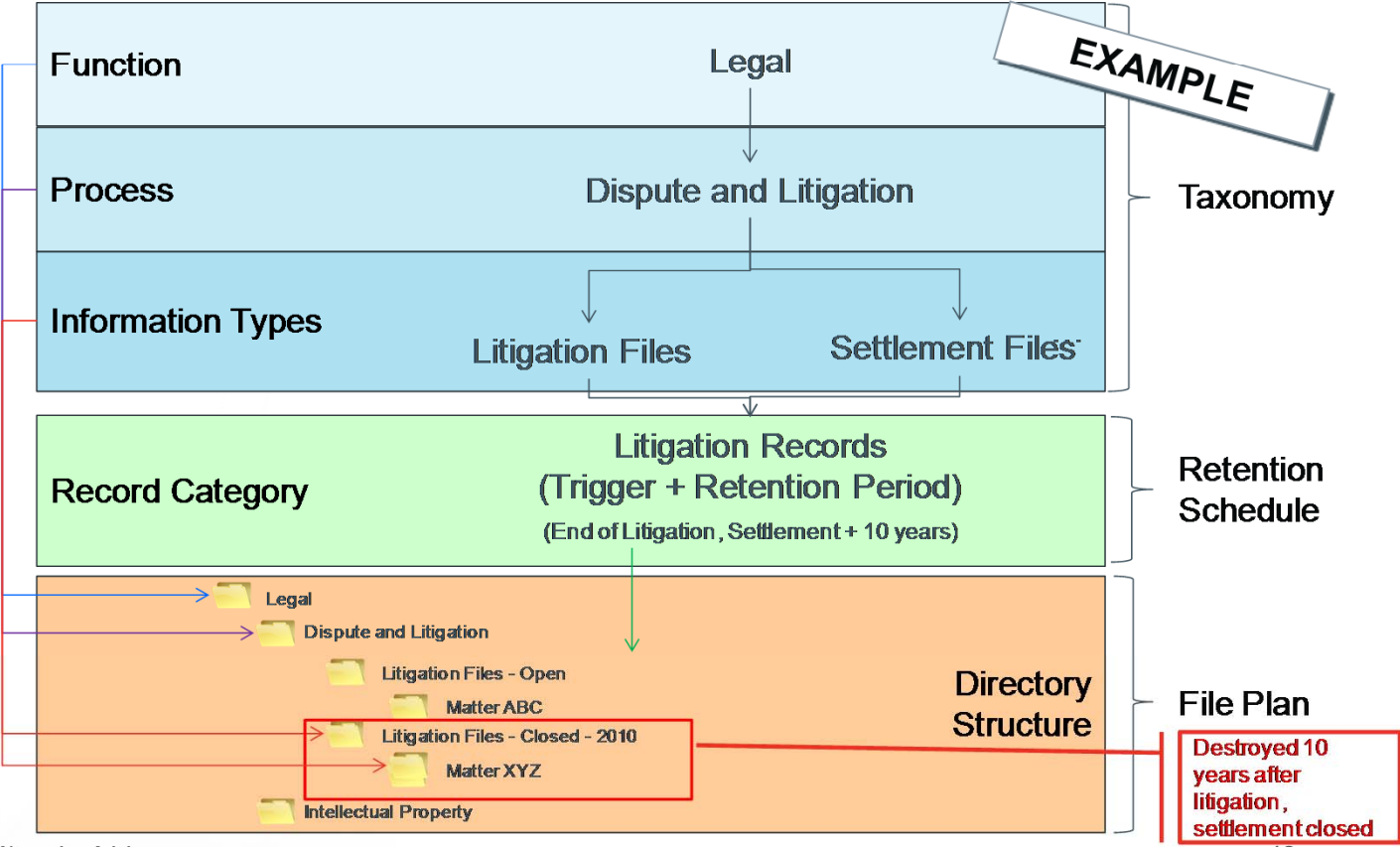
- Identify classification requirements
  - Review and refine
    - Conduct gap analysis to ensure all departments, functions across organization are included in the classification matrix
    - Identify any exceptions, contradictions or issues that may create challenges in developing the classification scheme



# Steps to Develop a Classification Scheme

- Create classification matrix
  - Level 1 – functions – *remember, a function is not about what department you're in, it's about what you do*
  - Level 2 – processes
  - Level 3 – information types

# Classification Scheme Applied



# Steps to Develop a Classification Scheme

Function	Process	Information Types
Logistics	Customer Service Management	Orders Placed, Processed, Confirmed
Logistics	Customs Management	Export Licenses, Export Training and Guidance, Export Licensing Determination, Export Control Files, Pertinent Information in Export Control Decision Process, Customs Declarations

**EXAMPLE**

Source: AIIM International

# Steps to Develop a Classification Scheme

- **Confirm and finalize classification scheme**
  - Review draft classification scheme with relevant POCs
    - POCs do not need to review the whole scheme – only those sections that apply to what they do
  - Confirm all functions, processes and information types are captured accurately

# Steps to Develop a Classification Scheme

- **Confirm and finalize classification scheme**
  - **Ensure language and terms used are concise, consistent and user-friendly**
    - **Establish agreed controlled vocabulary**
  - **Resolve any issues, as needed**
  - **Finalize and approve classification scheme**

# Consider Internal Organizational Environment

- Records and classifications exist ALL over the organization
- Can inventory by business process or by organizational function
- Need to take into account the classification schemes already in place.
- Key areas to start first:
  - Accounting
  - Human Resources
  - Regulated Operations
  - IT
  - Legal / Contracts
  - Purchasing

# Classify By Content Not Format

- Record Should be Unalterable in any format
- E-mail is NOT a Category
- Classify by Content and Context
  - Example:
    - Classification called contracts indexed by client and date
    - Records in classification should be format independent

# Information Analysis

- Organization Charts
- Present Retention Plans and Classifications
- Organizational Workflow and / or Business Process
- Understand that many departments may see the same information in different ways.



# Example of a Hierarchical Scheme



**Dept - Accounting**  
**Documents / Records**

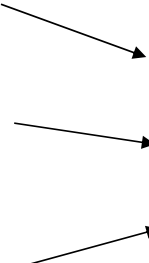
Invoice



Supporting  
Documentation

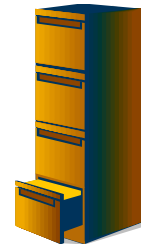


Correspondence

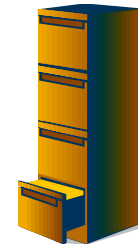


**Invoice  
Index**

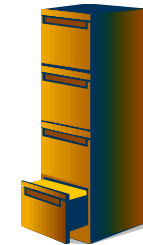
## Classifications



Fixed  
Assets



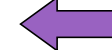
Accounts  
Receivable



Accounts  
Payable

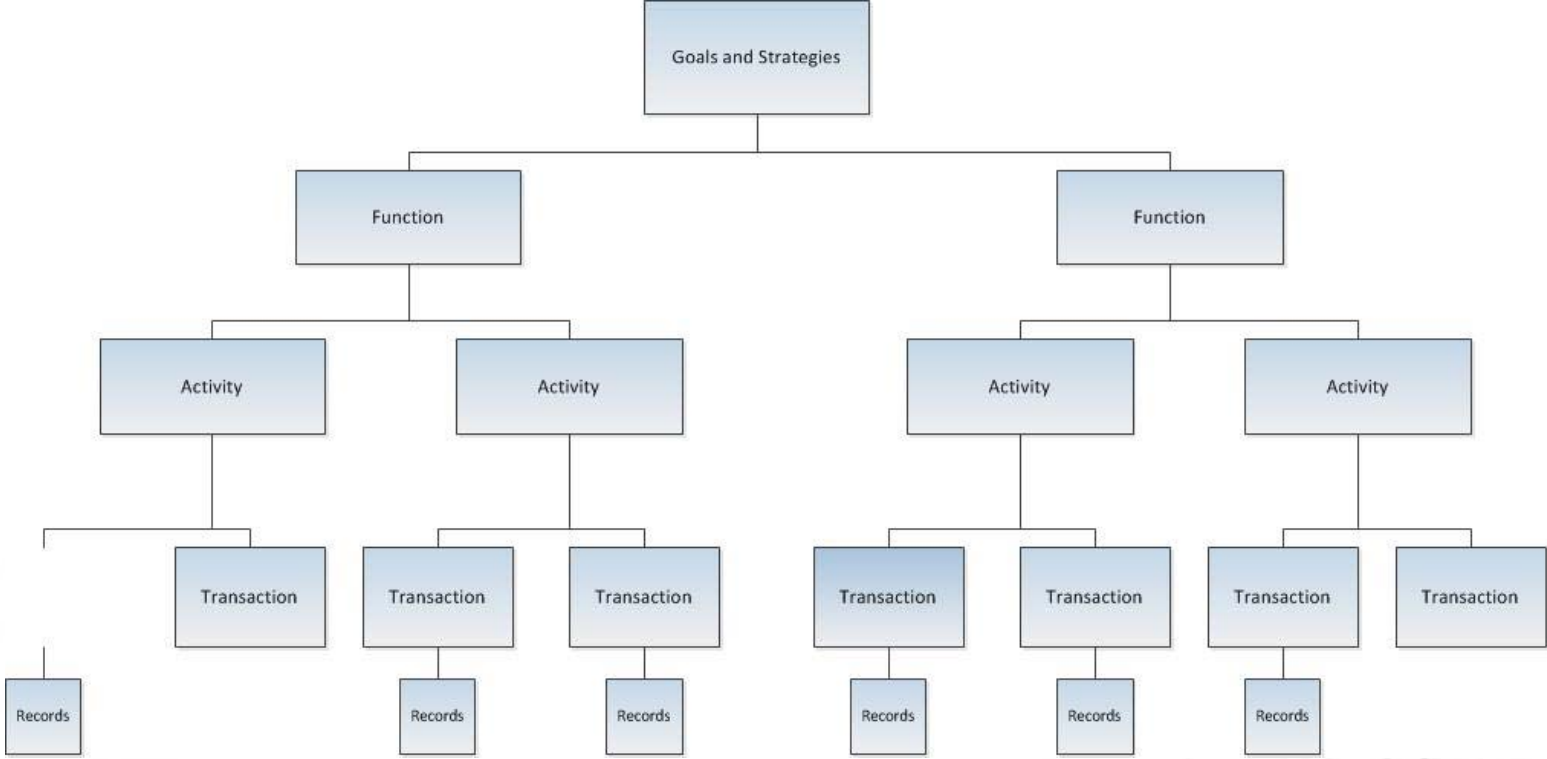


**Sub-Class**

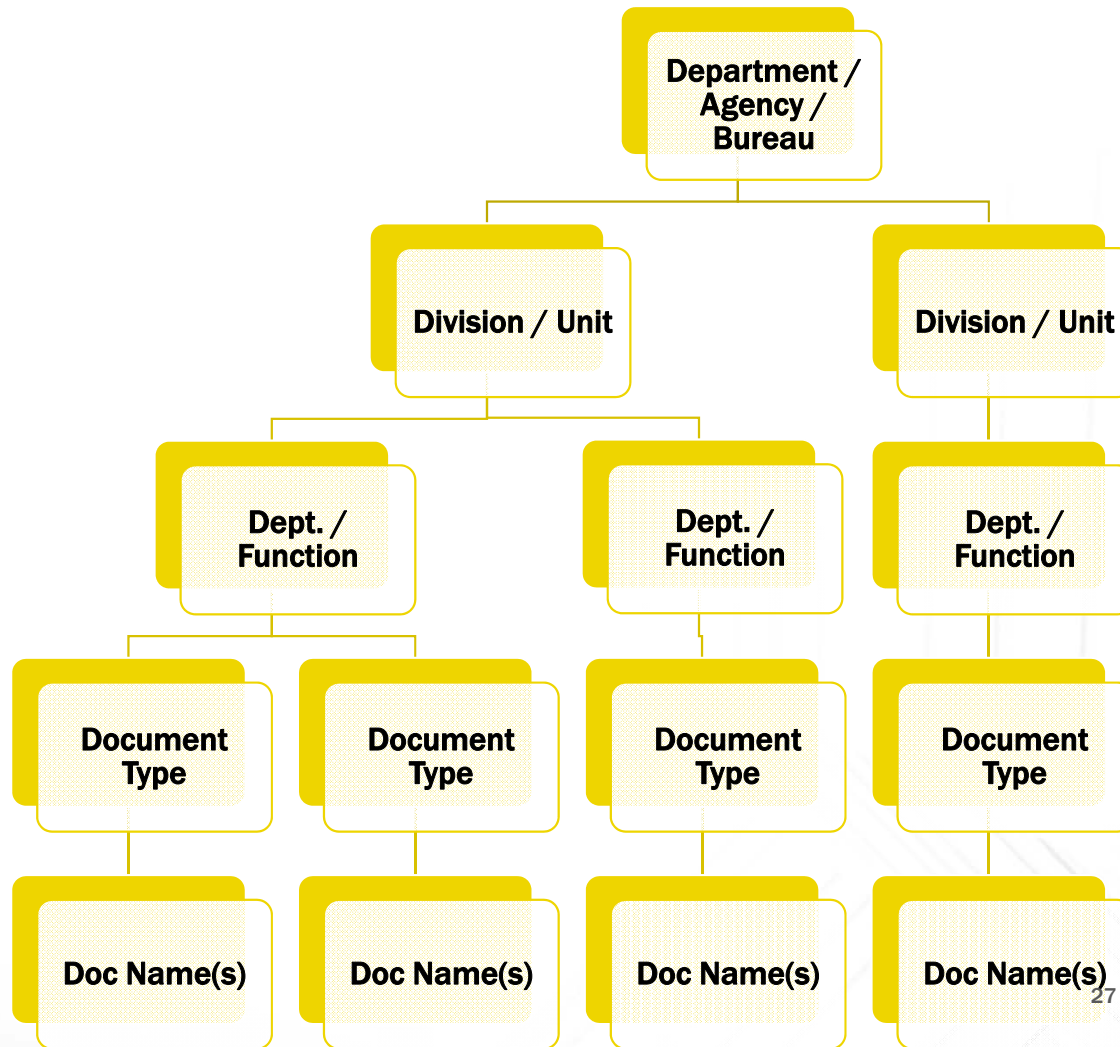


Invoices<sub>25</sub>

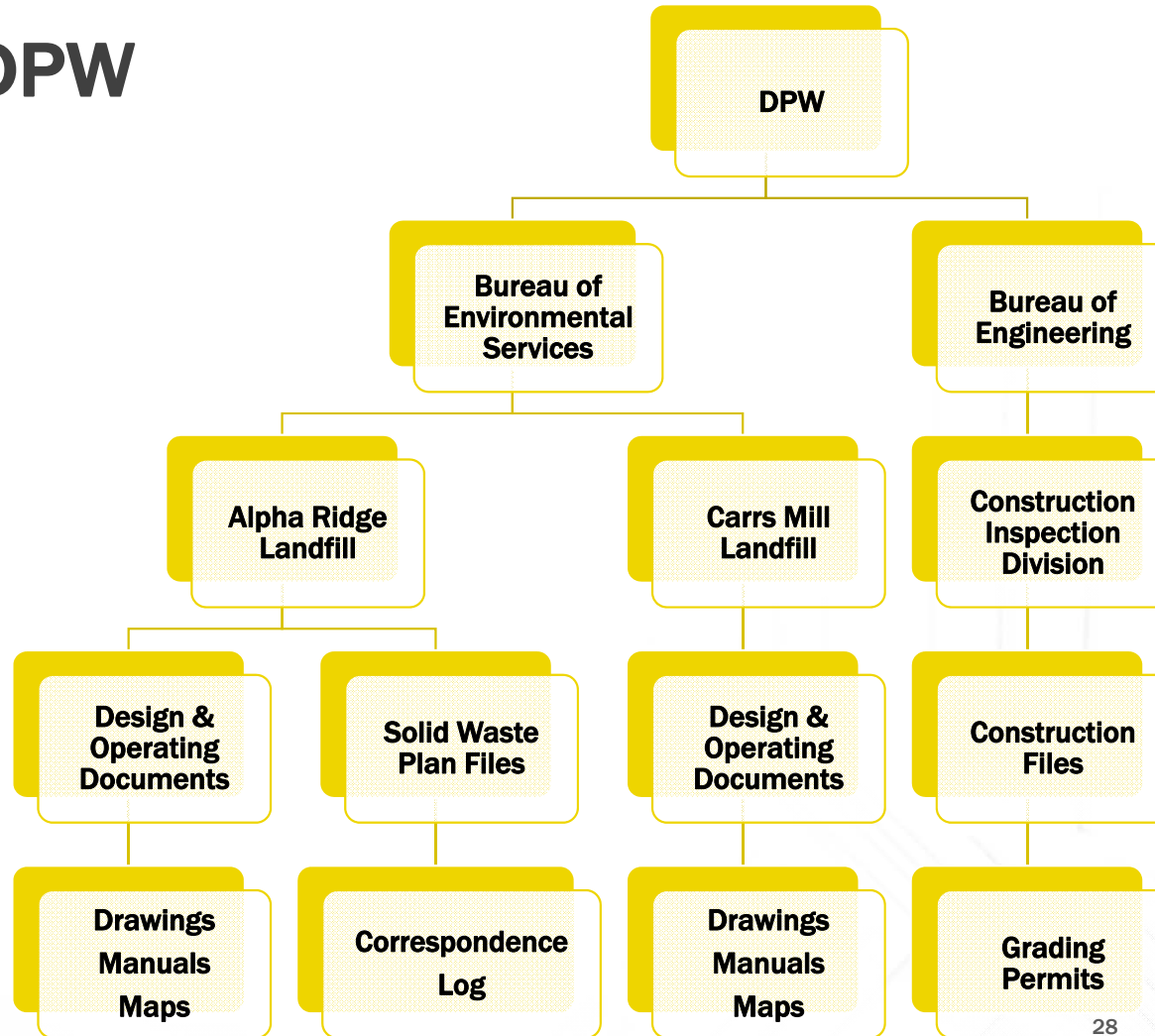
# Functional Classification Scheme



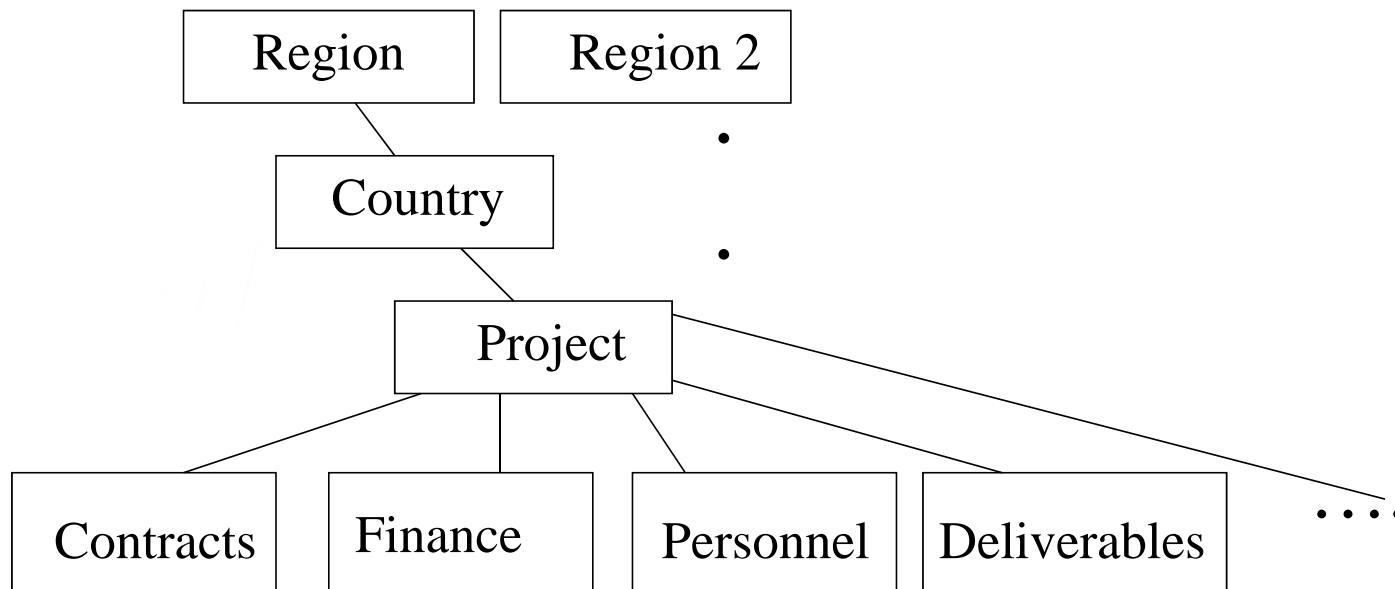
# Sample Tree



# Sample Tree DPW



# Example of Implied Classification Scheme



# Classification Scheme Should Be System / Environment Agnostic

- Windows-based Shared Drive
- SharePoint Repositories
- Paper Repositories ????
- New Document Management / Records Management Systems

# Controlled Vocabularies

- **Controlled vocabularies provide a way to organize knowledge for subsequent retrieval.**
- **Controlled vocabulary schemes mandate the use of predefined, authorized terms that have been preselected by the designer of the vocabulary**
- **Standard naming conventions for not just classifications**
  - **Document naming conventions**
  - **Indexing conventions**
  - **Metadata standards**

# Advantages of Controlled Vocabularies

- **Controlled vocabularies are often claimed to improve the accuracy of free text searching, such as to reduce irrelevant items in the retrieval list.**
- **The use of a controlled vocabulary can dramatically increase the performance of an information retrieval system**
- **Needs to be intuitive and not obstructive**



# Definition of Metadata

- Data about the Data
- Who, what, when, where, and how of an electronic document
  - Author, date created, last accessed, alteration dates, etc.
- Innate availability of imbedded metadata depends on the properties of the file type (i.e. MS Office, E-mail, WordPerfect, graphic files, Web addresses, etc.)

# Purpose of Metadata

- Enables the creation, registration, classification, access, preservation and disposition of records through time and within and across domains
- Used to identify, authenticate and contextualize records and the people, processes and systems that create, manage, maintain and use them and the policies that govern them

Source: Mary Naughton, Church Pension Group; and Lori Ashley, Tournesol Consulting, Metadata Standard for Records Lifecycle Management, 2011

# Types of Metadata

- **Descriptive** – information used to search and locate an object
- **Administrative** – technical information about the structure of computer systems and file formats
- **Structural** – description of how the components of the object are organized
  - **Rights management metadata**
  - **Preservation metadata**

Source: Mary Naughton, Church Pension Group; and Lori Ashley, Tournesol Consulting, Metadata Standard for Records Lifecycle Management, 2011

# Metadata in Litigation

- Without metadata a document / record is considered incomplete
- Provides information that is crucial to authenticating a document in court
- “Behind the scenes” quality
- “Digital fingerprints” can tell a story about computer user or history of a file
- Potential for losing metadata is enormous
- Potential for misusing metadata is also a challenge e.g. NSA

# Landmark Metadata Ruling

- *National Day Laborer Organizing Network v. U.S. Immigration and Customs Enforcement Agency*
  - Broke ground by its requirement that federal FOIA productions must be accompanied by metadata
  - Judge ordered government to produce metadata
  - Stressed criticality of cooperation and communication among counsel to reach early agreement on the form of production for pertinent ESI

Source: Mary Naughton, Church Pension Group; and Lori Ashley, Tournesol Consulting, *Metadata Standard for Records Lifecycle Management*, 2011

# Metadata Types

**There are two main types of file metadata that can be gathered during electronic processing.**

- **File system (Windows 8, Apple OS, etc.) information:** This information is stored independent of the file itself - the file system contains information about the file
  - File names, dates, path locations, sizes, etc.
- **Internal e-mail and file metadata:** This information is stored within the file itself and varies depending on the type of file in question
  - E-mail (To, From, Subject, etc.)
  - File (Author, body text, etc.)

# At Creation and Capture

- Level of aggregation
- Unique identifier
- Originator/Creator
- Owner
- Business Function
- Dates (open, close, cut-off)

Source: Mary Naughton, Church Pension Group; and Lori Ashley, Tournesol Consulting, Metadata Standard for Records Lifecycle Management, 2011

# Sample Metadata Elements

- Records class
- Format
- Creation date
- Creator
- Author
- Business Unit/ Support Group
- Document Type
- Publication Type
- Stakeholder
- Unique Identifiers
  - Document number
  - Customer number
  - Policy number
  - Invoice number

Source: Mary Naughton, Church Pension Group; and Lori Ashley, Tournesol Consulting, Metadata Standard for Records Lifecycle Management, 2011.



# Metadata Considerations

- **Information that can be:**
  - **Captured automatically at creation**
  - **Validated from other systems**
  - **Manually entered**
  - **Used across system platforms**
  - **Provide cross-functional value**
  - **Sustained over time**

# Risks of Losing or Modifying Metadata

- **When Accessing or Reopening files**
  - Last accessed date changes
  - E-mail Create Date made not be the correct date
- **Burning to CD or DVD**
  - Some software doesn't preserve the create and/or last accessed dates
- **Losing or truncating original path**
- **Forwarding e-mail messages**
- **Moving data between different operating systems**

# Classification With Respect to RM

- **Classification can be a method of assigning retention/disposition rules to records. Similar to the Declare function, this can be a completely manual process or automated, depending on the particular implementation.**
- **At a minimum, the user can be presented with a list of allowable file codes from a drop-down list (manual classification).**
- **Ideally, the desktop process/application can automate classification by triggering a file code selection from a property or characteristic of the process/application.**

# Records and Information Management (RIM)

- **File Plan:** A logical order of documents or files
- **Records Retention Schedule (RRS):** A document that lists the types of records maintained by each dept. of an organization, and specifies a period of time after which destruction is authorized as a matter of policy

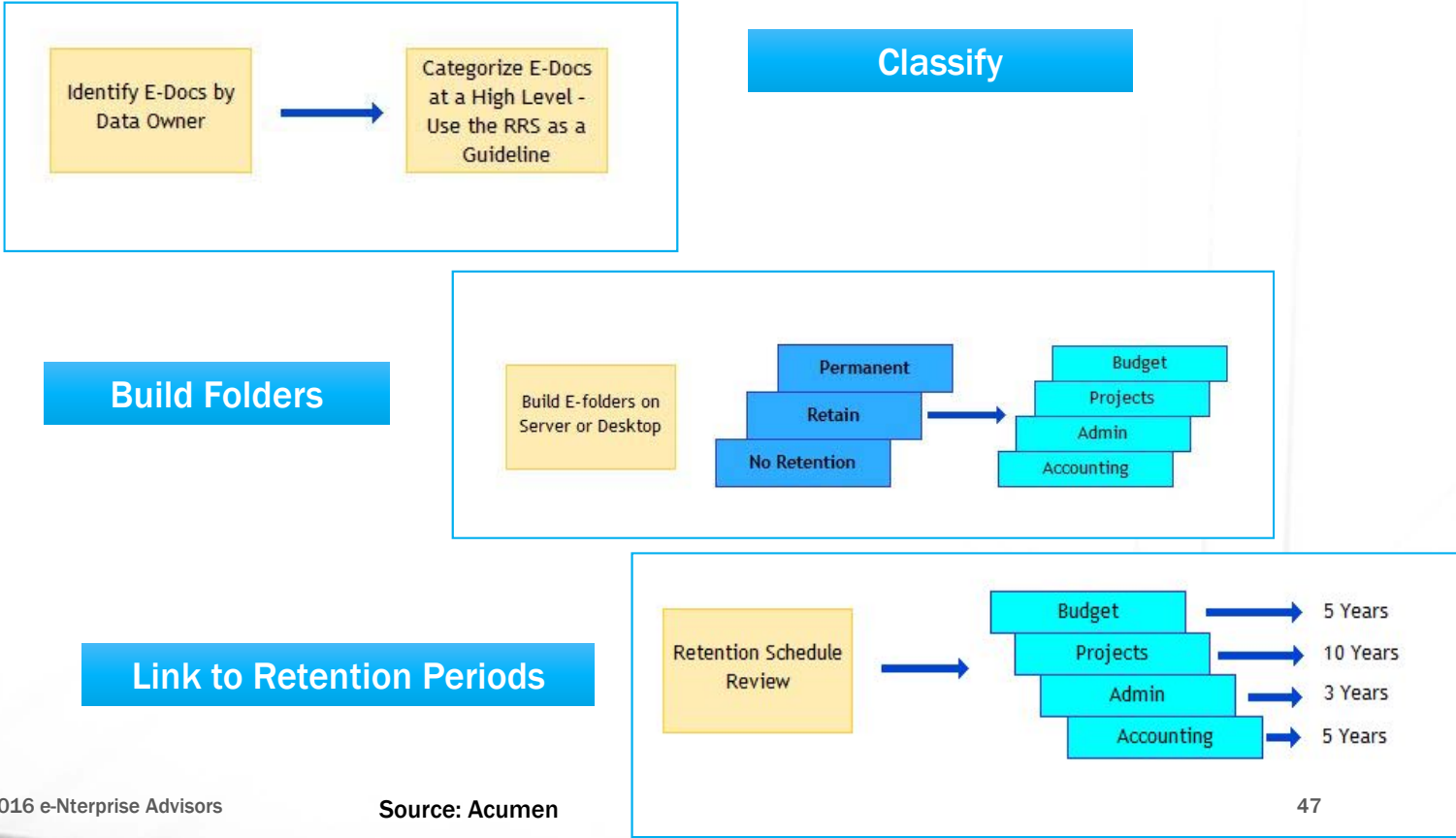
# Classification Plan Effectiveness

- **Classification systems should be designed to make sure that records are named consistently, allocate user permissions to particular types of records, distribute responsibility for management of each record type, identify disposition actions for records, and define media types for each record series.**
- **Classification systems that are designed on the basis of the business activity or function are more useful than those based on departmental names.**

# Key Questions To Ask

- Identify the most frequently retrieved records you use by title. For example; Permits, Contracts, Accounting Invoices etc.
- What data elements (key pieces of information) do you use to retrieve from these records? Example: Contract Number, Date, Name, Invoice Number etc.
- Where are the records you retrieve typically located within the office area?
- Of the record types you listed under question #1, please estimate the volume of records you retrieve on a monthly basis, by record type. Example: Invoices, 10 times per month.
- What are the most critically important records you use on your job?

# Electronic Documents Shouldn't Be Ignored

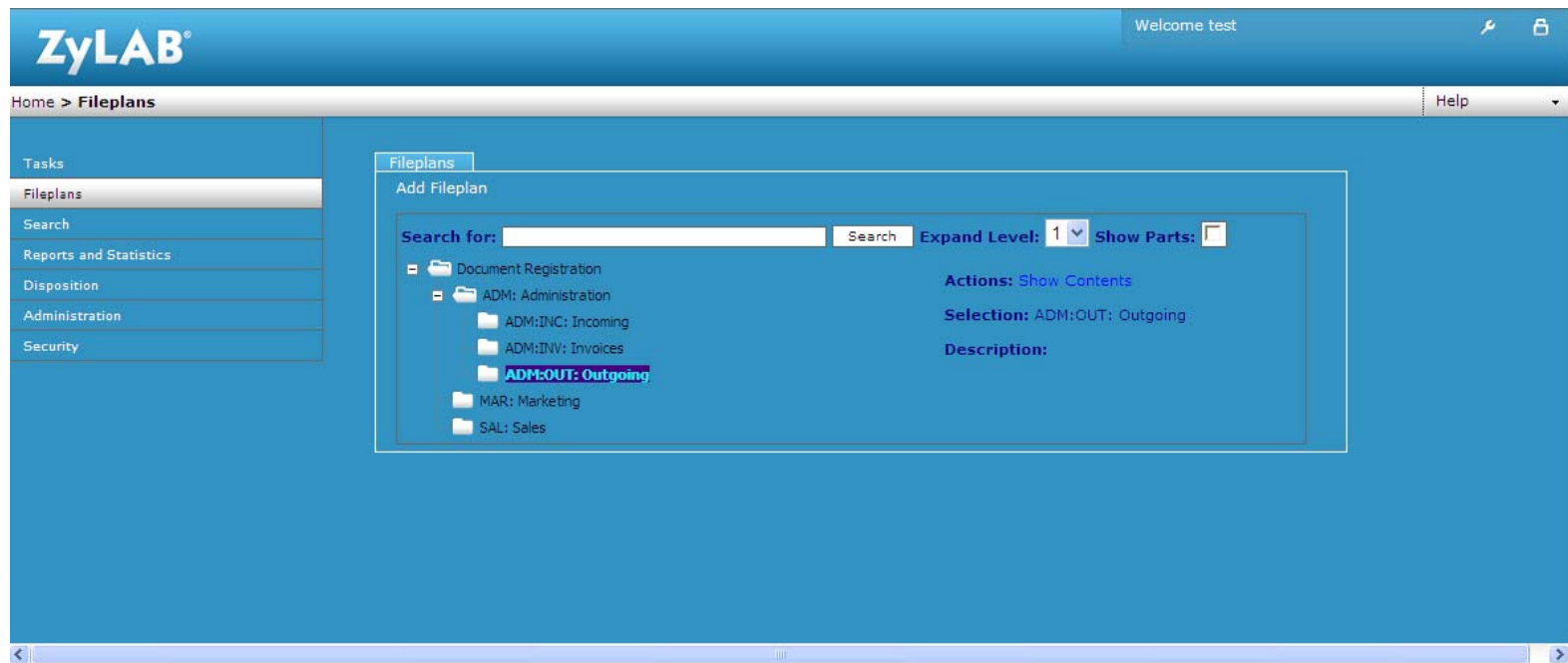


# Set Up a Filing Standard

- Using classification and index information gathered refer to Retention Schedule when creating Filing Plan
- Filing Plan
  - How records are organized
  - Who is responsible for doing what
  - When “clean-up” should occur (e.g. annual file purge)
  - What happens to the records when they are no longer needed in the office



# Electronic File Plans: Example



Source: ZyLAB, Inc.

# Additional Notes

- Provide cross referencing for paper and electronic folders/files
- Use the “Big Bucket” approach to classification if you can
- Create a map of file cabinets, if necessary
- A master control log may be required for large file stores

# e-Nterprise Advisors



Longwood, FL 32779 USA  
E-Mail: [pemery@ecmscope.com](mailto:pemery@ecmscope.com)  
[pemery@infogovgal.com](mailto:pemery@infogovgal.com)

[www.ecmscope.com](http://www.ecmscope.com)  
Twitter: @ecmscope

We provide "real world" advice to our clients on all the aspects associated with information governance and records management