A Practical, Hands-On Approach to Taxonomy Development

Monica Crocker, CRM, PMP, CIP

Greater Chattanooga Area ARMA

April 10, 2018

Together we'll go far

Speaker Introduction and Session Overview

- ■ Speaker
  - Certifications
  - Experience
  - Records philosophy

- ■ Session Overview
  - Define Taxonomy
  - Describe a Taxonomy Development approach
  - Discuss some "gotchas"

Why do you care about Taxonomy?

- Records are only useful if you can find them

- Full text searching doesn't work

- **Create standards: Consistency is critical to system and process integrity**

- Improve findability
  - Helps locate everything (Recall)
  - Helps eliminate irrelevant results (Precision)
  - Allows grouping of search results
  - Helps maintain organizational knowledge
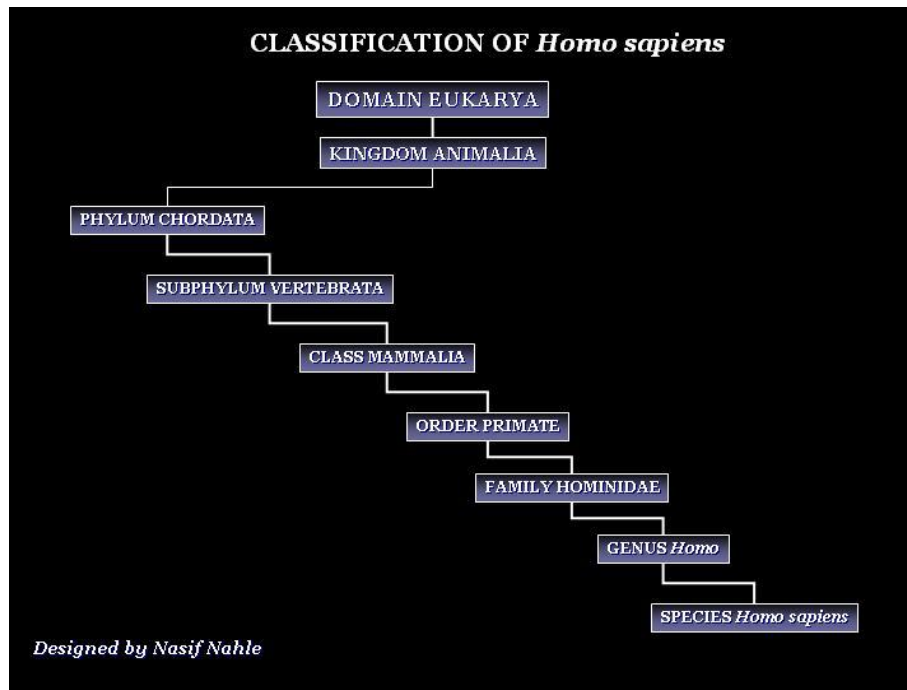 - Provides multiple ways to get to information

Further Benefits of Taxonomy Design

- Putting things in categories helps manage volume

- More electronic filing means different search tools needed

- Applies structure to unstructured content

- To help manage records lifecycle (such as deletion)

- Helps document compliance with regulations

- Individuals don't want to figure it out themselves

KEEP CALM AND Organize ON

Taxonomy, defined

- A system of classifying things based on their relationships

- A way to categorize content (classification)

- A way to get to content (navigation)

- Classifications go from more general to more specific:

CLASSIFICATION OF *Homo sapiens*

DOMAIN EUKARYA
KINGDOM ANIMALIA
PHYLUM CHORDATA
SUBPHYLUM VERTEBRATA
CLASS MAMMALIA
ORDER PRIMATE
FAMILY HOMINIDAE
GENUS *Homo*
SPECIES *Homo sapiens*

Designed by Nasif Nahle

Types of Taxonomies

- Most common types: subject, functional, organizational, hybrid

  - Subject taxonomies require a robust cross-reference of synonyms, preferred terms and related terms and don't support information security and retention requirements

  - Organizational taxonomies don't hold up to re-orgs or support cross-functional content and processes

  - Functional taxonomies support information security and retention requirements

  - A hybrid of the last two, with subject incorporated into metadata may be the best

Taxonomy Example
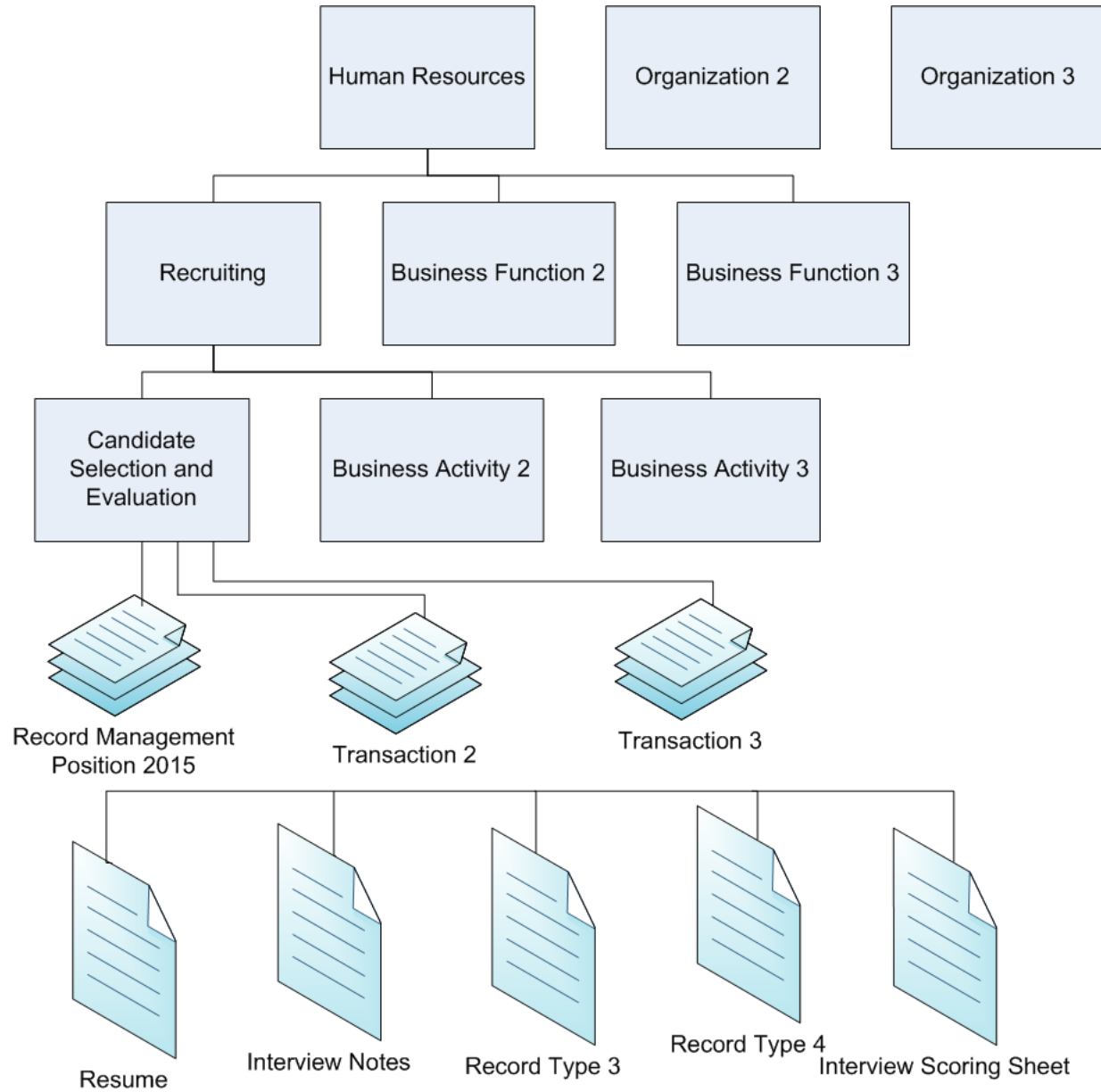
Exercise #1 Put the items on your list into groups and name the groups. And come up with a name for everything together.

Group #	Item	Group #	Item	Group #	Item
	Cleaver		Cheese grater		Peanut butter
	Garlic		Measuring cup		Footed cake plate
	Betty Crocker cookbook		Dish towel		Oven mitt
	Brandy		Can of beans		Cast iron skillet
	Cheesy snack		Margarine		Cream
	Drain opener		Crackers		Lettuce
	Tea bags		Mixer		Onions
	Plate		Recipe card box		Apron
	Milk		Cat food		Wooden spoon
	Sugar		Frozen spinach		Paring knife
	Can opener		Fresh spinach		Bottle of wine
	Sandwich bags		Spatula		Water filter pitcher
Title for All the Items listed:					

When do you need a Taxonomy Development project?

- When your organization is tired of wasting time looking for information

- When rolling out a new record/content/document management tool (you need a metadata model)

- Use cases

  ● With auto-classification

  (Network drive cleanup)

  ● Without auto-classification

  (Document Management, SharePoint)

The Relationship of Metadata

- Data about data, or in this case, records

- Descriptive fields that identify a record's characteristics

- Metadata model

  - Metadata fields assigned to all the items in a group

  - User entered or system generated

  - Controlled vocabulary

  - Inherited or automatic metadata values Example: Author, Title, Document Date, File Type, Subject, Category

- Can link items in separate buckets by a common value

- Dublin Core Metadata (ISO Standard 15836-2003): The Dublin Core Metadata Element Set is a vocabulary of document properties for use in resource description. Specific to records management.

Other Related Terminology

- Index values – the actual metadata values associated with a particular item

- Key words – user defined values in a keyword field or search terms used in a full text search

- Text mining – using a statistical inventory of the full contents of a library of items to define common terms

- Auto-classification – using analysis of the text and metadata associated with content to automatically assign index values to it, particularly classification values

Good Metadata

- Should describe the content itself, within the context of the taxonomy characteristics

- Should allow users to search with the information they already have

- Should be enough metadata to identify each item uniquely AND NO MORE

- Should be immutable (not change over the life of the item)

- Should be intuitive/understood by a novice
  - No acronyms
  - No jargon

More Metadata Criteria

- Each item in a controlled vocabulary should be mutually exclusive

- Makes electronic records legally admissible (proves trustworthiness)

- Intuitive indexing leads to improved system usability

- Automated indexing only works if both systems use similar values

- Controlled vocabulary terms should be long lived

- Indexing should be foolproof

- Clarity is key

Things to Avoid

- Folders

- Team member names (except in the Author field)

- Status fields

- Due dates (work triggers)

- Technology specific terms ("PDF" files)

- Duplicating data from a source system

- Too many fields

**DANGER**

MIGHT COUSE BRAIN TO EXPLODE

Exercise 2: Pick one of your groups and define fields you would use to find items in that group (maximum of 10)

Metadata Field Names	Sample Values (Controlled Vocabulary)
Name	*Cleaver, Cheese grater*

Check your Metadata Model

- Do your metadata fields:

  - Identify content and be able to distinguish it quickly from irrelevant content

  - Support business processes that center on the content

  - Support legal and regulatory compliance needs

  - Provide better search results

  - Secure the content
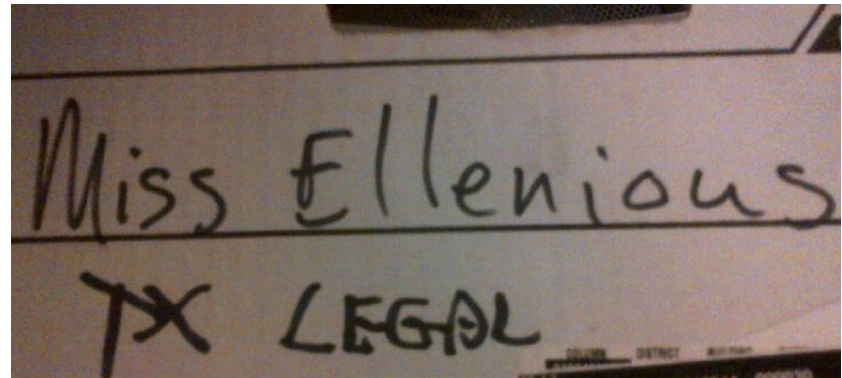
- How can you reduce the amount of manual data entry needed on those fields?

Taxonomy and Metadata Model Development Process

- Define scope…how will the taxonomy be used and by whom?
  - To find content
  - To put content on hold
  - To delete content
  - To audit a process
  - To identify policy violations
- Gather Subject Matter Experts
  - People who work with the records
  - People who manage the records
  - System gurus
  - Taxonomy experts
 - People who report on the work related to the records

Taxonomy and Metadata Model Development Process, cont.

- Start with your Business Classification Scheme or Functional Hierarchy

  - Business Classification Scheme: Level 1: Information Series, Level 2: Theme

  - Functional Hierarchy: Level 1: Business Function, Level 2: Business Activity: Level 3: Transaction

  - Add the relationships among items at each given level (is one a start and the next a finish or one a creator and the other a consumer)?

- Determine any system limitations (field formatting or length, dependent choice lists, length of choice list)

- Research existing (prebuilt) taxonomies available
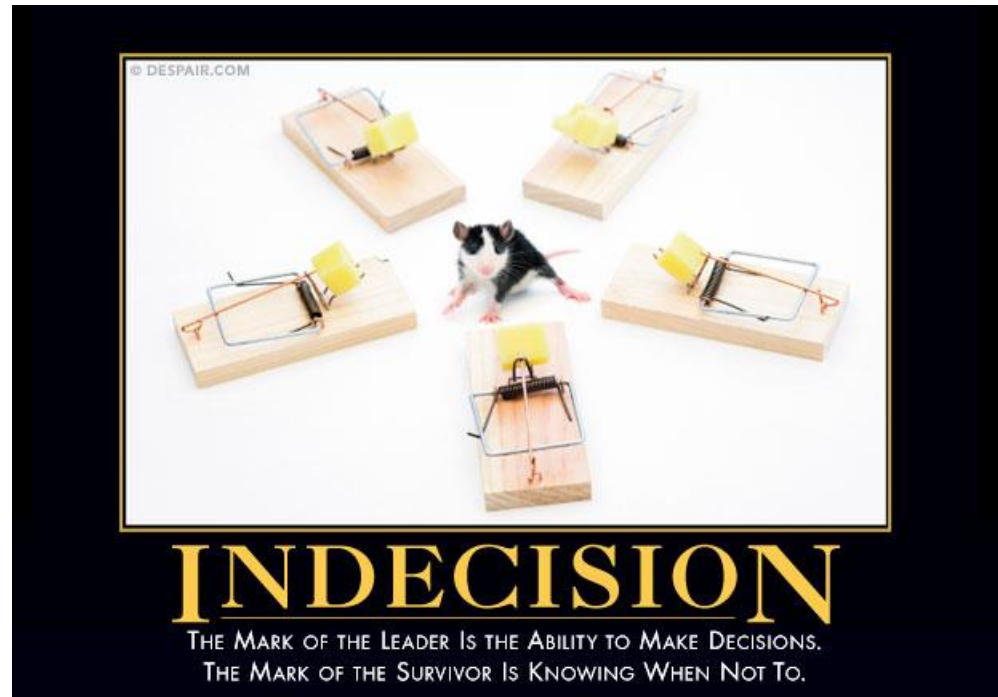
Process, cont.

- Collect all existing organization "taxonomies" including:

  - Records Retention Schedule

  - Organizational Charts

  - Security classifications

  - Off-site storage inventory

  - Budget

  - File plans

  - Business process documentation

  - Document or records inventory

UNSTOPPABLE
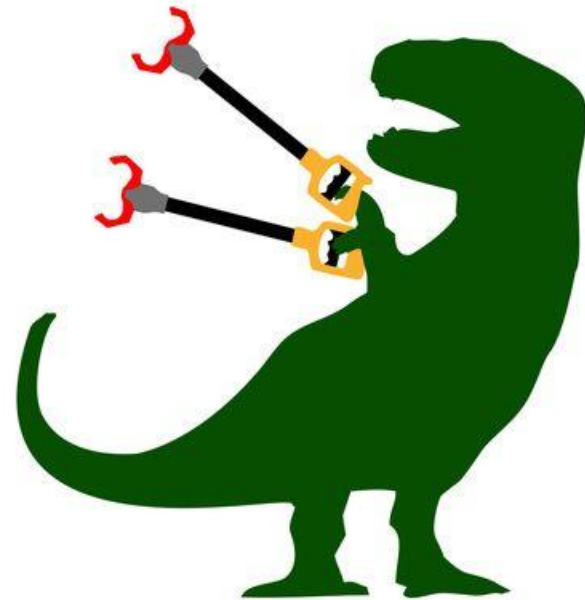
Process, cont.

- Develop taxonomy
  - Determine how it will be documented
  - Build a thesaurus (map preferred terms to potential terms)
  - Figure out how to handle items that don't fit – create new category or categorize as "other"
  - Decide if you will supplement it with folksonomy (keywords, comments) or social tagging
  - Define responsibilities

© DESPAIR.COM

INDECISION
THE MARK OF THE LEADER IS THE ABILITY TO MAKE DECISIONS.
THE MARK OF THE SURVIVOR IS KNOWING WHEN NOT TO.

Process, cont.

- Test taxonomy

  - Does the schema give participants in each business activity the information they need to do their jobs?

  - Does everything fit into the model?

  - Can you identify the appropriate retention period for all the items?

  - Conduct a meaningful user test

- Modify the taxonomy as appropriate

- Implement the taxonomy

  - Lots of training and communication

  - Quality assurance process

- Update and review periodically

  - Define a process for user suggested revisions

  - Define a process to evaluate and implement revisions

Exercise 3: : *Look at the groups you initially defined. Do you want to move anything to another group? Would the same metadata and user access settings apply to all the items in that group?*

Group Number	Group Name	Group Security – unique or same?	Metadata Fields – unique or same?	Move some items to another group?
	Exercise 1	**Exercise 3**		
1				
2				
3				
4				
5				

Conclusion

- It's just another huge, complex records management program/project with potential for significant impact on business productivity

- EXCEPT that it requires records management, business process and library sciences skills

- Questions?

- Speaker Contact Information

  - Monica Crocker, CRM, PMP, CIP

  - monica.crocker@wellsfargo.com

monica, dear, that was a precious little story. now, be a sweetheart and fix mommy another martini.