

# An Introduction to Enterprise Taxonomies

Heather Hedden  
Senior Taxonomy Analyst

SLA Conference, Philadelphia, June 15, 2011



# An Introduction to Enterprise Taxonomies

Heather Hedden  
Senior Taxonomy Analyst

SLA Conference, Philadelphia, June 15, 2011



- Taxonomy Consultant, Project Performance Corporation
- Continuing Education Instructor, “Taxonomies & Controlled Vocabularies,” Simmons College Graduate School of Library & Information Science
- Author, *The Accidental Taxonomist* (Information Today Inc., 2010, ([www.accidental-taxonomist.com](http://www.accidental-taxonomist.com)))
- Previously taxonomist with First Wind, Viziant Corporation, and Information Access Company/Thomson Gale (now Cengage Learning)

# About Project Performance Corporation

## Energy/Environment

Green strategies for government and industry:

- Air quality and climate change
  - Greenhouse gas reduction
    - Carbon management
- Environmental risk mitigation
- Environmental impacts of transport
- Information and data management

**1,200-person**

**multi-disciplinary team of scientific & technical experts**

- Scientific subject matter experts
- Systems engineers and architects
- Policy and regulatory specialists
- Project management professionals
- Certified Information technology experts
  - Security professionals

## Information Management

- Program and Project Management
  - Earned Value Management
  - Performance Measurement
- Program Assurance and Evaluation
  - Business Process Improvement
  - Security Policy and Compliance
  - Communications/Outreach and Facilitation

## Infrastructure

- Systems Engineering and Technical Assistance (SETA)
- Capability Maturity Model Integration (CMMI)
- Earned Value Management
- Configuration Management
- Technical and Advisory Support
- Independent Verification & Validation (IV&V)

## Enterprise Solutions

- Master Data Management and Data Governance
  - Business Intelligence
  - Adaptive Data Warehousing
  - Enterprise Architecture
- Infrastructure Systems Engineering
  - Knowledge Management
    - Portal Solutions
- Enterprise Content Management
- IT Optimization/Virtualization

1. Taxonomy Definitions
2. Enterprise Taxonomy Definitions & Characteristics
3. Taxonomies Compared with Thesauri
4. Enterprise Taxonomy Building
5. Enterprise Taxonomies and CMS or SharePoint
6. Enterprise Taxonomies and Search
7. Enterprise Taxonomy Issues

1. Taxonomy Definitions
2. Enterprise Taxonomy Definitions & Characteristics
3. Taxonomies Compared with Thesauri
4. Enterprise Taxonomy Building
5. Enterprise Taxonomies and CMS or SharePoint
6. Enterprise Taxonomies and Search
7. Enterprise Taxonomy Issues

## Broad designations:

- Controlled Vocabularies
- Knowledge Organization Systems
- Taxonomies

*Somewhat interchangeable*

## Specific types:

- Term Lists (pick lists)
- Synonym Rings
- Authority Files
- Hierarchical Taxonomies
- Faceted Taxonomies
- Thesauri
- Ontologies

*Increasing complexity*

## Controlled Vocabulary, Knowledge Organization System, Taxonomy

- An authoritative, restricted list of terms (words or phrases)
- Each term for a single unambiguous concept (synonyms/nonpreferred terms, as cross-references, may be included)
- May or may not have structured relationships between terms
- Policies/control for who, when, and how new terms can be added
- Supports categorizing/indexing/tagging/metadata management of content to facilitate retrieval/findability.

## **Controlled Vocabulary (CV)**

- commonly used libraries, museums, archives, database vendors, corporate digital asset management, records management
- Especially in indexing contexts

## **Knowledge Organization System (KOS)**

- commonly used in scholarly discourse, library school courses

## **Taxonomy**

- commonly used in corporate/enterprise applications and for public web sites

# Definitions: Specific Types

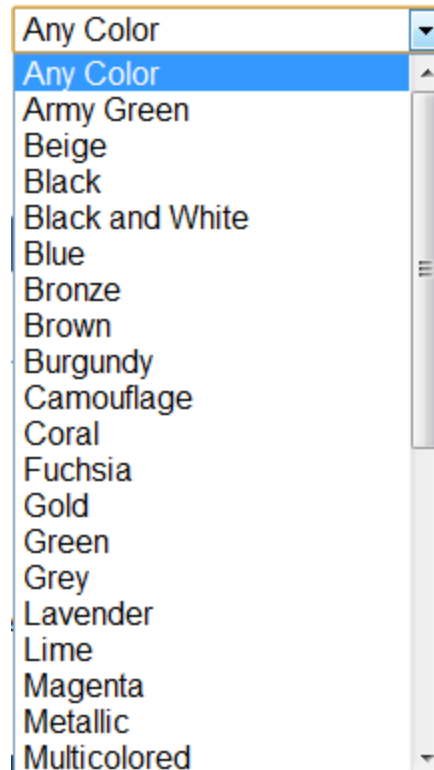
---

## Specific types:

- Term List
- Synonym Ring
- Authority File
- Hierarchical Taxonomy
- Faceted Taxonomy
- Thesaurus
- Ontology

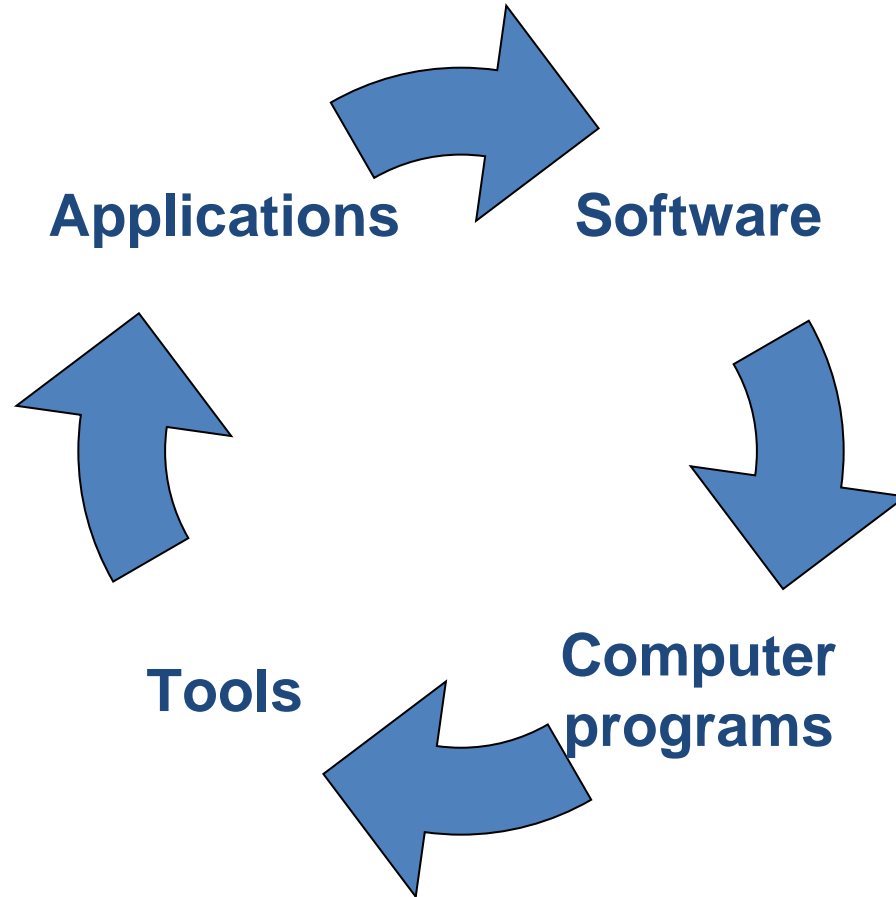
# Definitions: Specific Types

Term List - Flat pick list of terms without variants (nonpreferred terms)



## Definitions: Specific Types

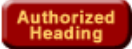
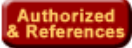
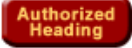


Synonym Ring - Has synonyms for concepts, but no “preferred” terms



# Definitions: Specific Types

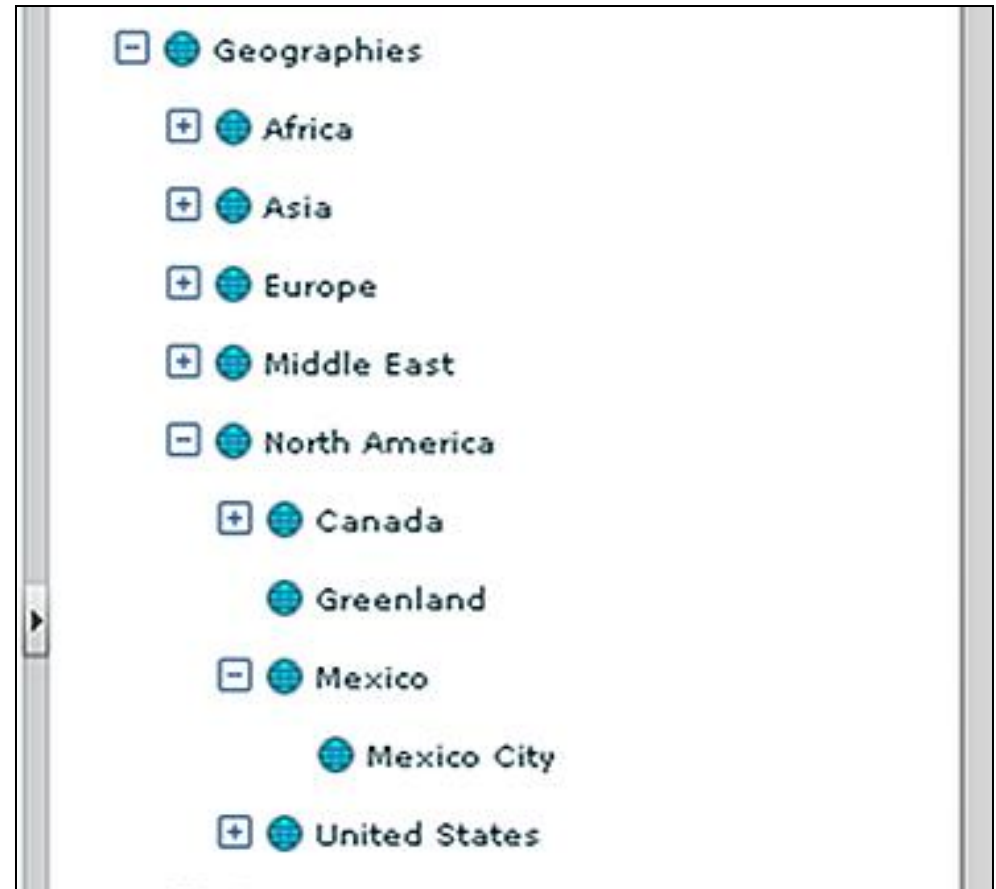
Authority File - Has nonpreferred and preferred terms, perhaps also an “authority” reference to source for each term

◀ Previous    Next ▶

#	Bib Records	<i>select icon in first column to...</i> View Authority Headings/References	<u>Type of Heading</u>
 1	0	Taxonomies of Organizational Behavior in Education Project	LC subject headings
 2	0	Taxonomists	LC subject headings
3	1	Taxonomists--British Columbia--Directories.	LC subject headings
4	2	Taxonomists--Directories.	LC subject headings
5	1	Taxonomists--Europe, Central--Directories.	LC subject headings
6	2	Taxonomists--Illinois--Directories.	LC subject headings
7	1	Taxonomists--India--Directories.	LC subject headings
8	1	Taxonomists--Mexico--Biography.	LC subject headings
 9	0	Taxonomy & ecology	LC subject headings
 10	0	Taxonomy and ecology	LC subject headings
 11	0	Taxonomy, Animal	LC subject headings

# Definitions: Specific Types

Hierarchical Taxonomy -  
Has broader  
term/narrower term  
relationships that  
include all terms to  
create a hierarchical  
structure



# Definitions: Specific Types

Faceted Taxonomy -  
Has sets of different types/aspects which the user selects in combination to refine a search by.

## Narrow Your Search

### + Search Within Results

#### – Locations Served

- Arizona
- Arkansas
- California - North
- California - South
- Colorado
- [+] More

### + Search Within # Miles

#### – Company Type

- Manufacturers
- Custom Manufacturers
- Distributors
- Service Companies
- Manufacturers' Reps

### + Certifications

#### – Ownership

- Minority-Owned
- Woman-Owned
- Veteran-Owned

## Definitions: Specific Types

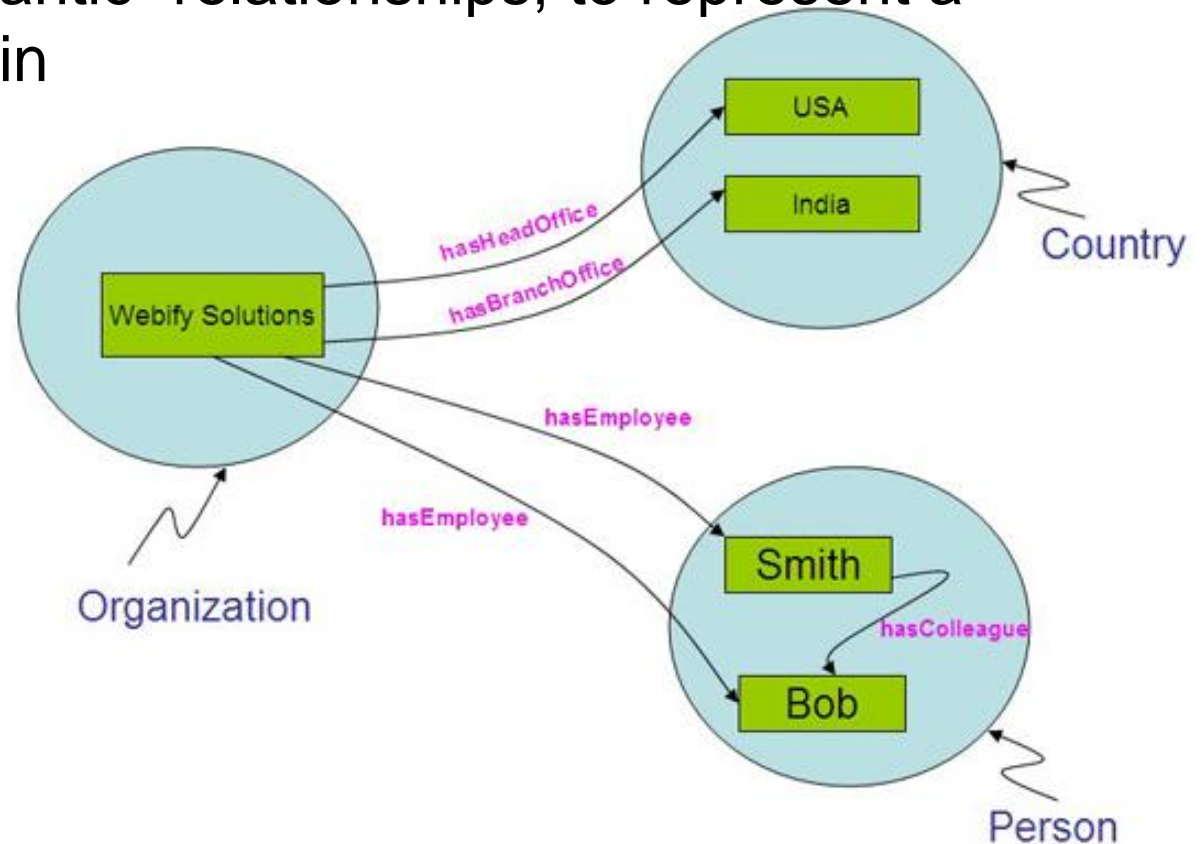
Thesaurus – Terms have hierarchical relationships, associative relationships, and nonpreferred/preferred terms.

### **materials acquisitions**

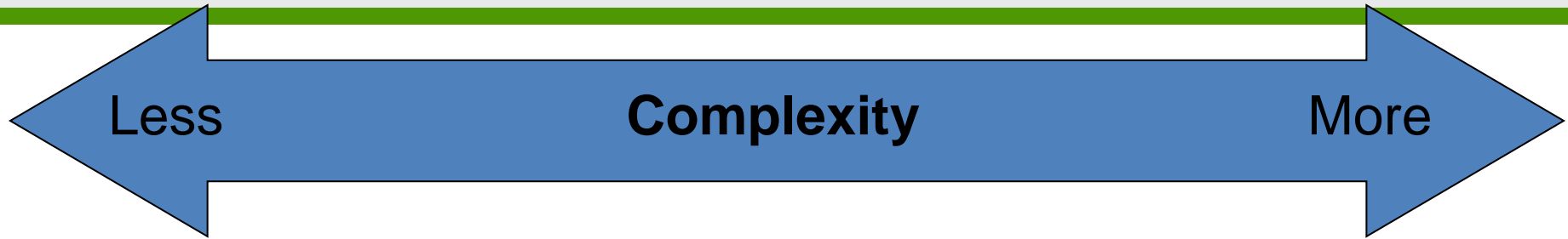
- UF acquisitions (of materials)  
library acquisitions
- BT collection development
- NT accessions  
approval plans  
gifts and exchanges  
materials claims  
materials orders  
subscriptions
- RT book vendors  
jobbers  
subscription agencies  
subscription cancellations

# Definitions: Specific Types

Ontology – Terms, depending on their assigned classes, have customized “semantic” relationships, to represent a knowledge domain



# Definitions: Specific Types



Term List/ Pick List	Synonym Ring	Authority File	(Hierarchical) Taxonomy	Thesaurus	Ontology
Ambiguity control	Synonym control	Ambiguity control Synonym control (preferred & non- preferred terms)	Ambiguity control Synonym control Hierarchical Relationships	Ambiguity control Synonym control Hierarchical Relationships Associative Relationships	Ambiguity control Synonym control Semantic Relationships Classes

1. Taxonomy Definitions
2. Enterprise Taxonomy Definitions & Characteristics
3. Taxonomies Compared with Thesauri
4. Enterprise Taxonomy Building
5. Enterprise Taxonomies and CMS or SharePoint
6. Enterprise Taxonomies and Search
7. Enterprise Taxonomy Issues

## Enterprise Taxonomies

1. As the broad definition of taxonomies
  - The name for any controlled vocabulary/knowledge organization system (in the broad sense) for use in an enterprise
2. As the more specific definition of taxonomies (especially as contrasted with thesauri)
  - A more specific kind of hierarchical and/or faceted taxonomy for use in an enterprise

- Taxonomy for *internal* use in addition or instead of external use
- Includes *all* of the enterprise, or almost all
  - All operating activities of an enterprise
  - Not necessarily general & administrative
- Often called a Business Taxonomy

# Characteristics of an Enterprise Taxonomy

- Custom-built, not purchased, taxonomy
- Not too deep or granular
  - to stay relevant for the entire enterprise
- Displays in an easy, user-friendly user interface
  - for use by *all* employees, not just information specialists
- Comprises a set of hierarchies/sub-taxonomies
  - which may or may not serve as facets
  - one “taxonomy” or multiple “taxonomies”?
- Implemented in an enterprise content management system (or any system being used as an ECM system)

1. Taxonomy Definitions
2. Enterprise Taxonomy Definitions & Characteristics
3. Taxonomies Compared with Thesauri
4. Enterprise Taxonomy Building
5. Enterprise Taxonomies and CMS or SharePoint
6. Enterprise Taxonomies and Search
7. Enterprise Taxonomy Issues

## Differences:

- Immediate purpose
- Focus
- Features
- Size
- Standards following
- Displays
- Common uses

## Immediate/direct purpose

- Thesaurus supports consistent indexing.
- Taxonomy supports end-user navigation/browsing/searching.

## Focus

- Thesaurus focuses on ***terms and their relationships.***
  - Think of Roget's thesaurus
  - Thus includes a term record display.
  - Hierarchies of just two terms are OK. Orphan terms may be OK.
- Taxonomy focuses on ***hierarchy and structure.***
  - Think of Linnaean taxonomy.
  - All terms are part of one or a limited number of large hierarchies.

## *Thesaurus*

### **Government lending**

- BT [Economic policy](#)
- NT [Veterans' loans](#)
- RT [Agricultural credit](#)
- RT [Federally-assisted loans](#)
- RT [Federally-guaranteed loans](#)
- RT [Government and business](#)
- RT [Government insurance](#)
- RT [Loans](#)
- RT [Student loan funds](#)
- UF American domestic economic assistance
- UF Federal aid to depressed areas
- UF Federal credit programs
- UF Federal domestic assistance programs
- UF Government loans

## *Taxonomy*

### National Policy

- . Economy Policy
- .. Government Lending
- ... Veterans' Loans

## Size

- Thesauri contain as many terms as needed to index the content in sufficient detail.
  - A single thesaurus can have hundreds or thousands of terms
  - Bigger is often better.
  - Growth does not necessarily impact structure.
- Taxonomies can become too difficult to use if they are too big
  - A taxonomy is usually broken into sub-taxonomies, hierarchies, with only dozens or a couple hundred terms each
  - Bigger is less often better
  - Growth often impacts the structure

## Standards

- Thesauri should comply with ANIS/NISO Z 39.19 especially with respect to various relationships between terms.
- Taxonomy need only comply with the definition of hierarchical relationships, and even then there is more flexibility.

For example, making related terms appear as narrower, such as:

- Companies under industries
- Members under association names
- Accessories under core products

## Display Formats

- Thesauri are:
  - Usually alphabetical
  - Sometimes hierarchical, as a view option (toggle)
  - Other display formats may be possible (top term, KWIC, KWOC)
  - Also has term record display feature
- Taxonomies are either:
  - Hierarchical
  - Faceted
  - Sometimes both (at different levels, not different views)
  - No other options; No term record display

1. Taxonomy Definitions
2. Enterprise Taxonomy Definitions & Characteristics
3. Taxonomies Compared with Thesauri
4. Enterprise Taxonomy Building
5. Enterprise Taxonomies and CMS or SharePoint
6. Enterprise Taxonomies and Search
7. Enterprise Taxonomy Issues

1. Build a new enterprise taxonomy from scratch
2. Create an enterprise taxonomy by combining existing department taxonomies (integrating or merging)
3. Both

## 1. Build a new taxonomy

- Engage stakeholders from different departments of the enterprise
  - Present a taxonomy workshop with stakeholders to increase their understanding and stimulate their ideas for contributing
  - Interview stakeholders to gain insight into their content management methods, content findability needs, and current difficulties
  - Conduct card-sorting exercises
- Conduct content inventory (high level) and sample content audit
- Identify potential hierarchies/types/facets

## 2. Integrate Existing Taxonomies

Combine separate (departmental) taxonomies into a single, larger taxonomy for combined use by all.

- Combines related, but not redundant taxonomies
- Taxonomies supplement each other in the same area
- Increases (multiplies) the number of terms
- May involve issues of interoperability, if legacy taxonomies were maintained in different systems
- Taxonomist deals with integrating structures more than integrating individual terms.

## 3. Create a new taxonomy and merge in existing taxonomies

Combine redundant taxonomies in same subject area into one

- The new enterprise taxonomy is designated the dominant/primary taxonomy into which to merge the other
- Compares two taxonomies side-by-side, term-by-term
- Taxonomist deals with terms more than structure

No matter how designed, follow up with:

- Taxonomy validation test/exercise
- Governance and maintenance plan

1. Taxonomy Definitions
2. Taxonomies Compared with Thesauri
3. Enterprise Taxonomy Definitions & Characteristics
4. Enterprise Taxonomy Building
5. Enterprise Taxonomies and CMS or SharePoint
6. Enterprise Taxonomies and Search
7. Enterprise Taxonomy Issues

Consider whether and how the content management system (CMS) will support:

- Dynamic faceted browsing
- Polyhierarchies
- Nonpreferred terms (via inputting and searching)
- Associative relationships

Consider how the CMS supports of manual and/or automated indexing of content with taxonomy terms

- Examine/design the manual indexing user interface

Consider where to maintain the taxonomy or how to import

- Does the CMS have an adequate taxonomy management feature?
- If not, where to maintain the taxonomy, and when and how shall the taxonomy be ported?
  - As CSV
  - As XML (or variants, such as ZThes)
  - As RDF

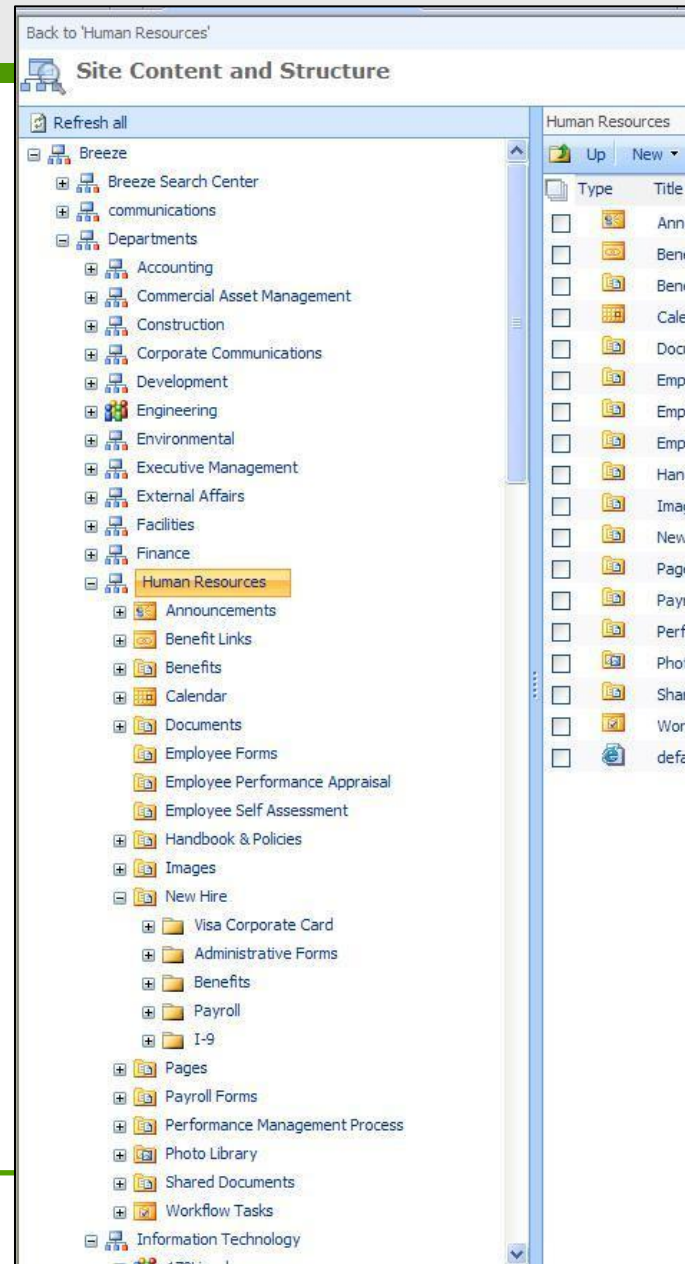
Taxonomies may be utilized in SharePoint:

- In the form of managed metadata to classify documents
- In the hierarchical naming and arranging of libraries and folders
- In support of search (with third-party search tools)

Taxonomy management for managed metadata is far superior in SP 2010 than 2007.

# Enterprise Taxonomy and SharePoint

- Don't overlook folder hierarchy design, though.
- Also, consider how to support search.



1. Taxonomy Definitions
2. Enterprise Taxonomy Definitions & Characteristics
3. Taxonomies Compared with Thesauri
4. Enterprise Taxonomy Building
5. Enterprise Taxonomies and CMS or SharePoint
6. Enterprise Taxonomies and Search
7. Enterprise Taxonomy Issues

## Not all search systems utilize taxonomies

- Many utilize:
  - variant lists of common misspellings
  - stored commonly used search strings
- These are *not* taxonomies/controlled vocabularies because:
  - They do *not* bring together synonyms/variants for *concepts*
  - No one *controls* the vocabulary
- Just because there is a term list, does not make it a controlled vocabulary/taxonomy
- If search software does not come with a taxonomy, a taxonomy can still be added/utilized.

Design taxonomies for search by configuring the search engine to look at the taxonomy and:

- Support all possible metadata fields with terms from controlled vocabularies
- Create nonpreferred terms to help match end-user queries and text strings
- Utilize an autocategorization, entity extraction, text mining, text analytics, etc.

1. Taxonomy Definitions
2. Enterprise Taxonomy Definitions & Characteristics
3. Taxonomies Compared with Thesauri
4. Enterprise Taxonomy Building
5. Enterprise Taxonomies and CMS or SharePoint
6. Enterprise Taxonomies and Search
7. Enterprise Taxonomy Issues

- Making the business case
- Getting “buy-in” and participation from all stakeholders
- Competing interests from different groups
- One taxonomy vs. multiple taxonomies
- Need to know the user/use cases
- Specific taxonomy term demands
- Commitment to taxonomy governance and maintenance

## Naming it

Across the entire enterprise “taxonomy” might not resonate

- Too technical, complex, intrusive, different, unfriendly, etc.
- TaxoCoP list post suggestions:
  - "Enhancing Search Experience"
  - "Increasing Search Results"
  - “Information Availability”
  - "Information Value Enhancement“
  - “Business Classification and Description”

Heather Hedden  
Senior Taxonomy Analyst  
Project Performance Corporation  
Heather.Hedden@ppc.com  
703-462-3746  
Corporate office: McLean, VA