



Taxonomies 101

Kathryn Breininger
and
Mary Whittaker
Boeing Library and Learning Center Services

October 6, 2006

Agenda

- Welcome and Introductions
- Goals and Objectives
- Part 1
 - What is the problem?
 - What are controlled vocabularies?
- Part 2
 - Controlled vocabulary standards
- Part 3
 - Development of a Taxonomy

Agenda

- Part 4
 - Maintenance & Metrics
 - Engagement & Governance
 - Applications & Interoperability
 - Strategy & Implementation
 - Benefits & Best practices
- Part 5
 - Boeing Case Study

Welcome and Introductions

- Introductions

- Name

- Library

- Experience building/using taxonomies

- What do you want to get out of this class?

- Logistics

Goals of this Workshop

- Review different types of controlled vocabularies
- Learn more about building a taxonomy
- Understand processes involved
- Learn about taxonomy applications
- Become familiar with ANSI standard
- Gain some practical hands-on experience
- Leave with some best practices

Part 1

What is the problem?

History

Aristotle “Father of Logic” developed taxonomies to organize his knowledge of the world.

What is the Problem?


- Information overload
 - Fortune 500 companies lost \$12 billion in 2000 “due to an inability to locate knowledge resources” (IDC, 2000)
 - “The amount of corporate data generated per employee doubles every 18 months” (NewsEdge, 2001)
 - Survey results show 42% of respondents spend up to 8 hours a week searching for information (Delphi, 2004)
 - Corporate knowledge workers spend an average of 7 hours a week searching for information, and another 5.6 hours a week analyzing it (Outsell, May 2006)
- Users don't know what terms to use for searching
- Individuals use multiple ways of organizing

Types of Content

- Product brochures
- Technical reports
- Word documents
- HTML pages
- Presentations
- Journal articles
- Video
- Images
- Books
- Still photos

The Problem with Full Text

- There is no control over terminology
 - Spelling errors
 - Must search every synonymous term
 - Cannot browse through concepts
 - Miss information on the topic of interest
 - Cannot associate concepts
 - Different definitions for concepts
 - No context for the concept



What are Controlled Vocabularies?

What is a Controlled Vocabulary?

- Represents concepts
- Controls spelling variants or errors
- Disambiguates homographs
- Eliminates ambiguity
- Merges duplicates and plurals
- Gathers synonymous terms
- Provides consistency
- Establishes relationships

Purpose of Controlled Vocabularies

- Translation
- Consistency
- Indicate relationships
- Browsible hierarchy
- Navigation
- Retrieval
 - Improved precision and recall

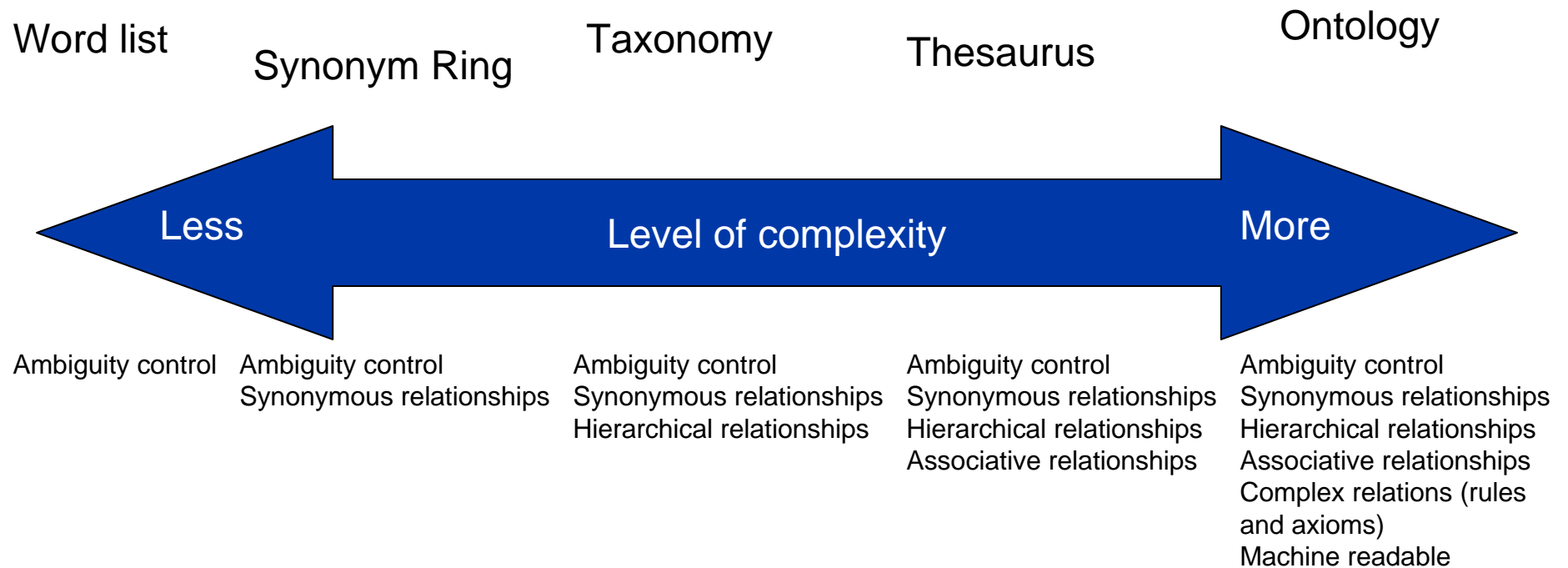
Value of Controlled Vocabularies

- Provides consistent terminology
- Provides consistent definitions for concept “translations”
- Shows relationships among terms
- Labels for browsing capabilities
- Improves search and discovery of relevant information
- Provides pick lists for content description
- Improves information discovery for reuse
- Organizes information
- Organizes content on web sites, file shares, desktops

Types of Controlled Vocabularies

- Word lists (controlled and uncontrolled)
- Synonym rings
- Taxonomies
- Thesauri
- Ontologies

Spectrum of Controlled Vocabularies



Based on ANSI
Z39.19-2005

What are Word Lists?

- Uncontrolled
 - Anything goes

- Controlled
 - Terms are predetermined

Exercise

- Create a grocery shopping list (by self)
- Work in groups and create a controlled list

Word Lists

- Used to display small sets of terms to be used for narrowly defined purposes
 - Drop down lists
 - Web menu items
- Pros
 - Simple to implement and use
- Cons
 - Provide no guidance for the user

Examples of Word Lists

- Uncontrolled
 - Mary calls it: **cat food**
 - Kathryn calls it: **kitten chow**

- Controlled
 - Safeway calls it: **pet food**

What is a Synonym Ring?

- A group of terms that are considered equivalent for the purposes of retrieval
- List of words with synonymous meanings
- Used for pulling in additional words for the same concept

Uses of a Synonym Ring

- With full-text searching
- Intranet search of web sites
- Behind-the-scenes activity

Synonym Ring Pros and Cons

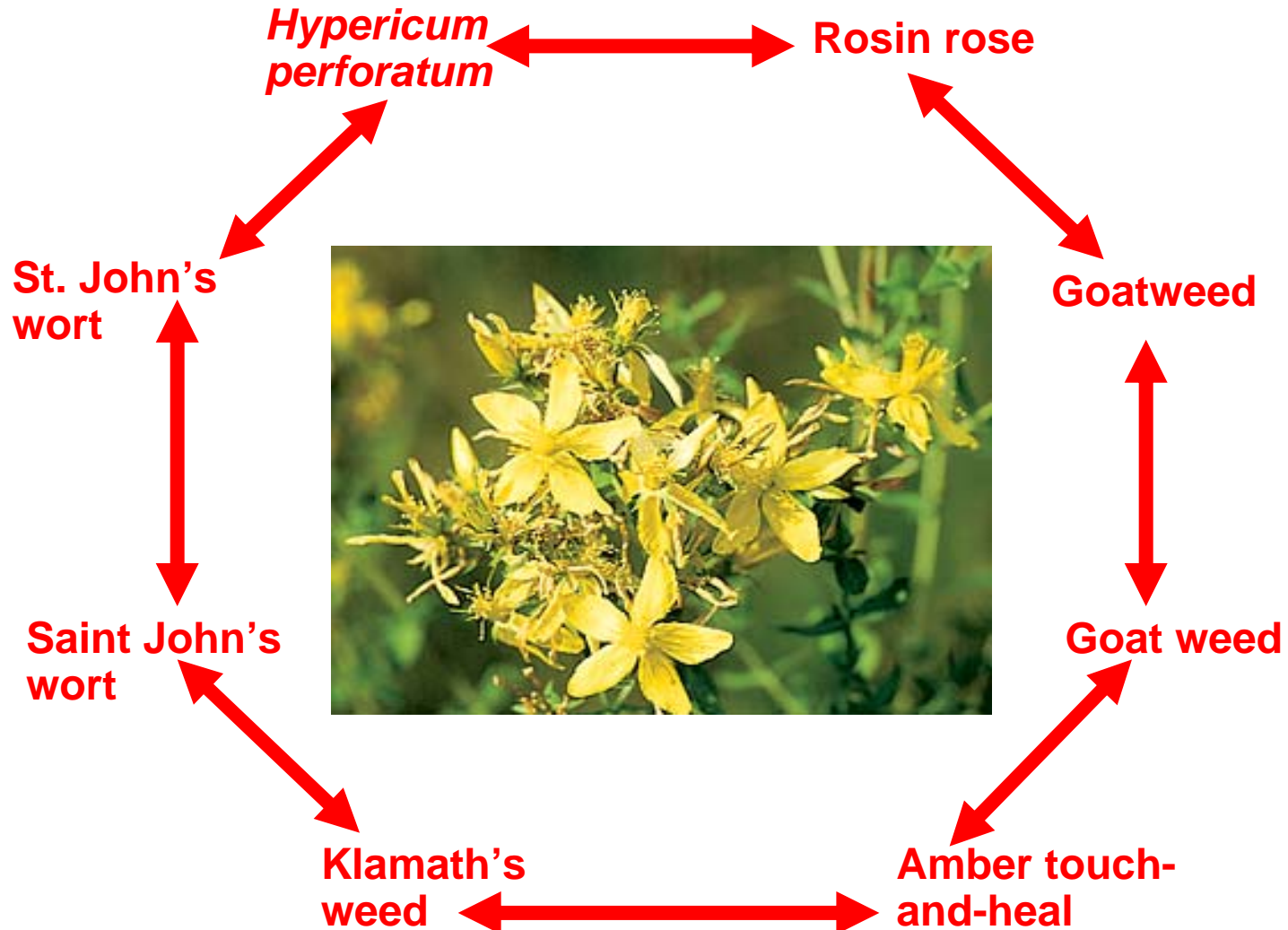
■ Pros

- Retrieves results on words that user may not have thought of to search on
- Synonyms treated equally in searching

■ Cons

- Manually created – time consuming
- Not used in indexing

Synonym Ring Example



Exercise

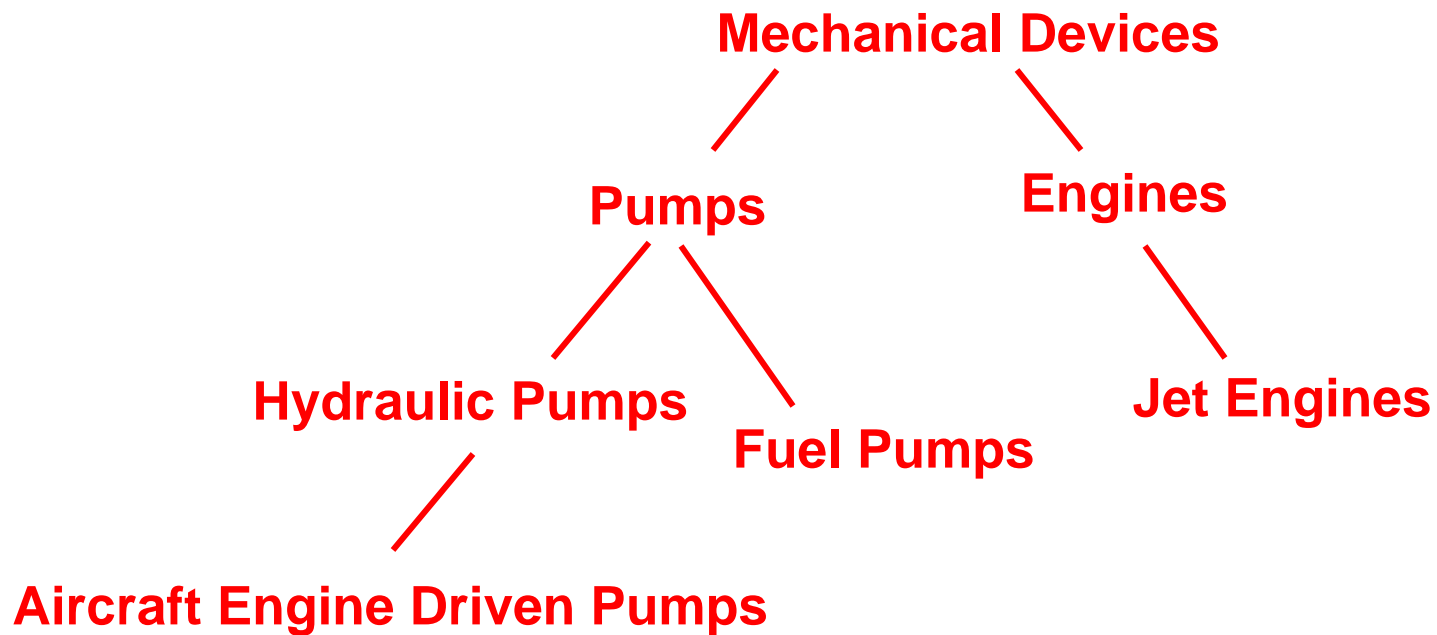
Pick a term from your controlled list and make a synonym ring (use your terms and any others)

What is a Taxonomy?

- A controlled vocabulary
- Relationships are one step broader or narrower
 - Genus/Species
 - Instance
 - Whole/part
- May include equivalent relationships
- Browsable hierarchical structure
- Used to consistently categorize information
- Provides search terms for the user

Taxonomy Relationships

— = Broader Term



Uses of a Taxonomy

- To consistently categorize information
- To provide browsable search terms
- To provide navigational structure for web sites
- For indexing applications
- For web navigation
- To lead users to the most specific terms in a particular domain

Taxonomy Pros and Cons

■ Pros

- Shows hierarchical relationships among terms
- Used by indexers and searchers who need appropriate specific terms

■ Cons

- Does not indicate additional types of relationships

Types of Taxonomies

- Descriptive taxonomies
- Navigation taxonomies

Siberian Tiger

Kingdom: *Animalia*

Phylum: *Chordata (or vertebrata)*

Class: *Mammalia*

Order: *Carnivora*

Family: *Felidae (cat)*

SubFamily: *Pantherinae*

Genus: *Panthera tigris*

Species: *Panthera tigris altaicia*

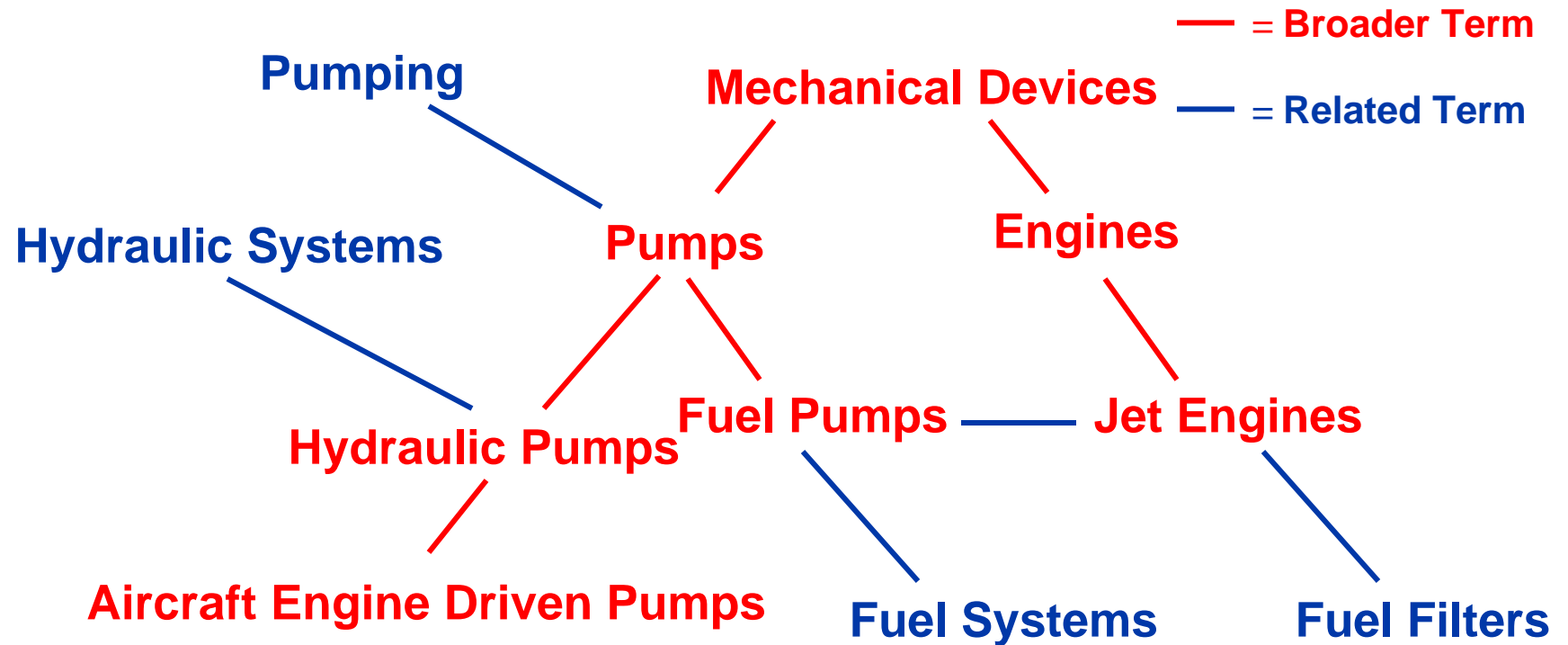
What is a Thesaurus?

- A controlled vocabulary
- Relationships are hierarchical, associative, and synonymous
- Presented in alphabetical order
- Used to index information, usually with multiple concepts

Thesaurus - Types of Relationships

- Hierarchical relationships
- Associative relationships
- Synonymous relationships

Thesaurus Relationships



Uses of a Thesaurus

- For indexing and searching applications
- Provides rich structure and cross references
- Can be narrow in scope for a limited domain or broad in scope and widely applicable
- Provides user search terms

Thesaurus Pros and Cons

■ Pros

- Information about various relationships among terms
- Entry vocabulary helps users find correct term
- Useful to both indexers and searchers needing specific terms

■ Cons

- Labor intensive
- Issues of complexity

DTIC Thesaurus

The screenshot shows a Microsoft Internet Explorer browser window titled "Verity K2 Document Display - Microsoft Internet Explorer". The address bar is empty. The menu bar includes "File", "Edit", "View", "Favorites", "Tools", and "Help". The page content features the DTIC logo (a blue oval with "DTIC" in yellow) and the text "Public STINET" in a blue bar. Below the logo, there are two links: "New Descriptor Search" and "Insert this Descriptor into a TR Guided Search". The main heading is "DTIC Thesaurus Entry". The entry details are as follows:

- Descriptor:** ALUMINUM ALLOYS
- Broader Terms:** [ALLOYS](#)
- Narrower Terms:**
- Used Alone For:**
- Use:**
- Used in Combination For:**
- Use in Combination:**

At the bottom of the page, there are two circular logos: the Department of Defense logo on the left and the DTIC logo on the right.

From: <http://stinet.dtic.mil/str/thesaurus.html>

National Library of Medicine - Medical Subject Headings

[Return to Entry Page](#)

MeSH Heading	Jet Lag Syndrome
Tree Number	C10.281.440
Tree Number	C10.886.425.200.500
Tree Number	F03.870.400.200.500
Scope Note	A chronobiologic disorder resulting from rapid travel across a number of time zones, characterized by insomnia or hypersomnolence, fatigue, behavioral symptoms, headaches, and gastrointestinal disturbances. (From Cooper, Sleep, 1994, pp593-8)
Entry Term	Time Zone Change Syndrome
Entry Term	Time Zone Syndrome
See Also	Travel
Allowable Qualifiers	BL CF CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SU TH UR US VE VI
Previous Indexing	Sleep Disorders (1974-1999)
History Note	2000
Unique ID	D020179

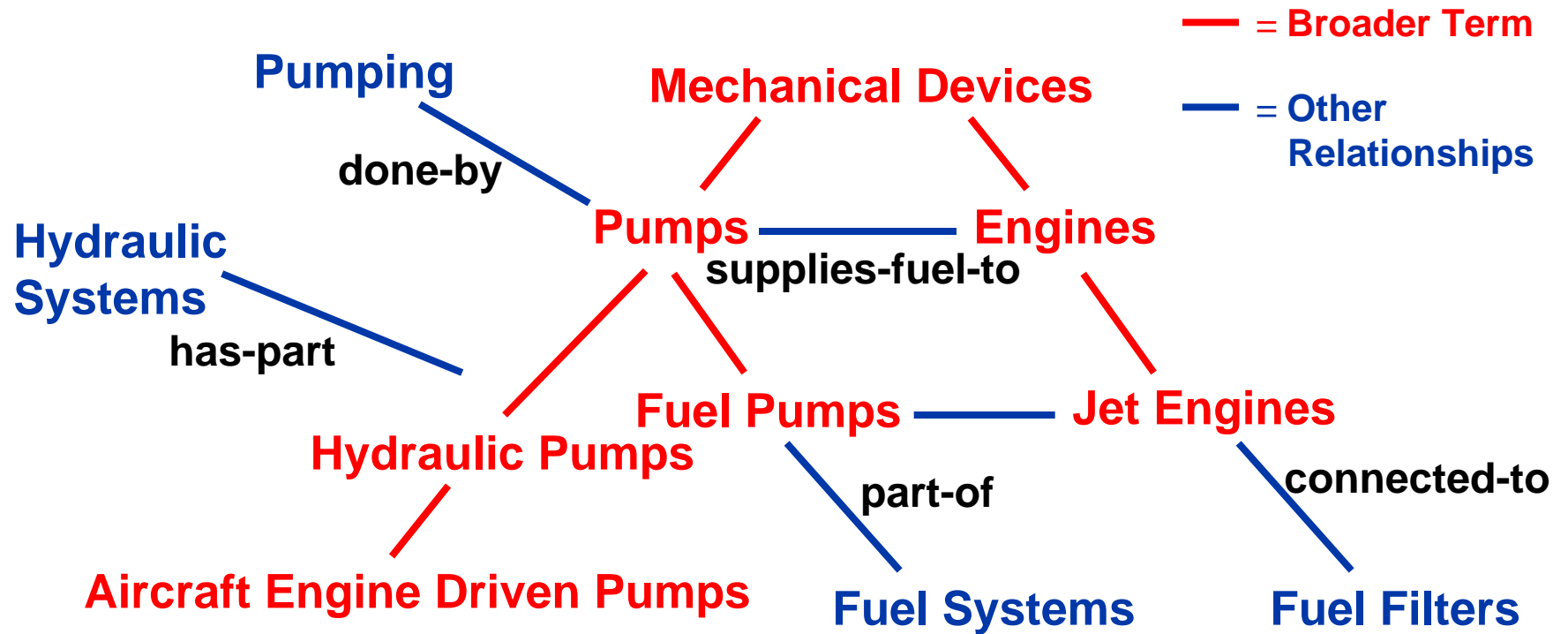
2006 MeSH - MeSH Descriptor Data

From: <http://www.nlm.nih.gov/mesh/MBrowser.html>

What is an Ontology?

- An explicit representation of a shared understanding of concepts
- A description of the concepts and relationships
- A set of definitions for a formal vocabulary of terms and relationships
- Agreement to use terms in a specific way (ontological commitment)

Ontology - Formal Relationships



Uses of an Ontology

- Knowledge sharing systems
- Interoperability of systems
- Search functions and services
- Concept-based structuring of information
- Intelligence analysis
- Pattern identification
- Semantic Web

Ontology Pros and Cons

- Pros

- Ontologies will be useful for machine agent searching

- Cons

- Relationships of ontologies are unique
- Difficult to map ontology to ontology

Other Types of Controlled Vocabularies

- Folksonomies
- Stackonomies
- Collabularies

Part 2

Standards

ANSI Z39.19-2005

“...presents guidelines and conventions for the contents, display, construction, testing, maintenance, and management of controlled vocabularies.”

-- Scope section of the standard

ANSI Z39.19 History

- Originally published in 1974
- Revised in 2005 (4th edition)
 - Expanded scope
 - Expanded audience
 - Expanded examples

Term Considerations

- Information space
- Validation of terms
- Term relationships
- Subject authorities
- Handling of candidate terms
- Specificity
- Unassigned terms

Scope of Terms

- Restrict terms to the information space
- Disambiguation issues relating to homographs

Homographs

- One of two or more words that have the same spelling, but different meaning
 - Earth
 - earth
- Try to avoid in the selection of terms for your taxonomy
- If needed, use a qualifier to clarify the term
 - mules (shoes)
 - mules (animals)

Qualifiers

- A word or words used with the term to make the meaning unambiguous, added after the term and enclosed in parentheses
 - elevators (lifts)
 - elevators (control surfaces)
- Use compound terms instead of single-words with a parenthetical qualifier
- May be omitted when a term is used in only one of its meanings within a taxonomy

Format of Terms

- Term should represent a single concept
 - Single-word **airplanes**
 - Compound term **military aircraft**

Types of Concepts

Things and their physical parts

houses, chimneys

Materials

silk, hydrogen

Activities or processes

ballroom dancing, calcification

Events

anniversaries, International Peace Day

Properties or states

meditation, tensile impact strength

Disciplines or subject fields

librarianship, psychology

Units of measurement

nautical miles, pounds

Proper nouns

International Space Station, Tahiti

Language specificity

- Process
 - publishing, indexing, cooking
- Agent
 - publishers, indexers, cooks
- Product
 - publications, indexes, meals

Grammatical Format of Terms

■ Nouns

- Preferably a noun or a noun phrase
 - **life**
- Gerunds or verbal nouns
 - **farming**
- Pre-modified noun phrase
 - **fiction writers**
- Post-modified noun phrase
 - **veteran's hospitals**
- Prepositional nouns
 - **burden of proof**

Grammatical Format of Terms

■ Adjectives

- Use sparingly
- Okay to use in a nominal way
- Good for pre-coordinated indexing or post-coordinated searching
 - Color or size examples: “red” or “large”

■ Adverbs

- Rarely used
- Okay in music domain
- Acceptable when in a concept with specialized meaning
 - **Very Large Scale Integration**

Articles

- Use when required
 - If meaning would be different without it, use an initial article
 - Part of a proper noun
 - Los Angeles
 - If meaning would be clear without it, do not use an initial article
 - Big Easy (not The Big Easy)
 - Okay to remove it from a title (if the title were a concept in a taxonomy)

Nouns

- Two types of nouns:
 - Count
 - You can answer the question “How many?”
 - Should be entered in a plural format
 - churches
 - Exceptions! Body parts, museum cataloging
 - lung, Stickley chair
 - Non-count
 - You can answer the question “How much?”
 - Should be entered in a singular format
 - electricity
 - Exceptions! If the term is already considered plural
 - ceramics

Singular Noun Usage

- Non-count nouns
- Abstract concepts
- Unique entities

spring water

peace

Taj Mahal

Singular and Plural Usages

- If the term in the plural represents a different concept from the singular, use both
- Indicate the distinction with a qualifier
 - marble (material)
 - marbles (game)

Selecting the Preferred Term

- Usage
- Spelling
- Abbreviations
- Acronyms
- Initialisms

Selecting the Preferred Form

- Use the full form
 - When the abbreviated form is not widely used
 - **Driving Under the Influence**, not **DUI**
 - When the acronym stands for more than one concept
 - FAR
 - **Federal Acquisitions Regulation**
 - **Federal Aviation Regulation**

Candidate Terms

- Slang
- Jargon
- Neologisms

cops

selective dissemination of
information

relationships

Other Nouns

- Popular names
 - St. John's wort
- Scientific names
 - Hypericum perforatum
- Loanwords
 - kindergarten
- Foreign language equivalents
 - Modus operandi = Method of work

Proper Nouns

- Variant forms cause problems
 - Mark Twain
 - Samuel Clemens
- Possible solutions
 - Include in a taxonomy
 - Enter into authority file
 - Leave as is

Compound Terms

- Expresses a single concept
 - **bowling alleys**
- Can be arranged in a genus-species relationship within a hierarchy
 - **gowns**
 - **wedding gowns**
- Has a focus (head noun) and a modifier (difference)
 - **border guards**
- Use natural language word order
 - **statements of work**

Compound Terms Guidelines

- Loss of meaning eating disorders
- Nondistinctive elements home economics
- Distinct meanings tennis shoes
- Metaphorical modifiers dogbone joints
- Adjectival modifiers artificial intelligence
- Proper names Reynolds number
- Common use information technology

Capitalization

- Use lower case
 - management
- Use initial caps for proper nouns, trade names, and acronyms
 - William Boeing
 - IBM Corporation
 - NATO
- Transcribe as seen (mixed case with unusual use of capitalization)
 - easyJet Airline Ltd.
 - eXtensible Markup Language

Punctuation

- Avoid parentheses, except when used with qualifiers
- Do not use hyphens, unless by not using them, it would lead to confusion
- Use apostrophes for possessives
- Use diacritical marks if they are required for a proper name
- Beware of using symbols!

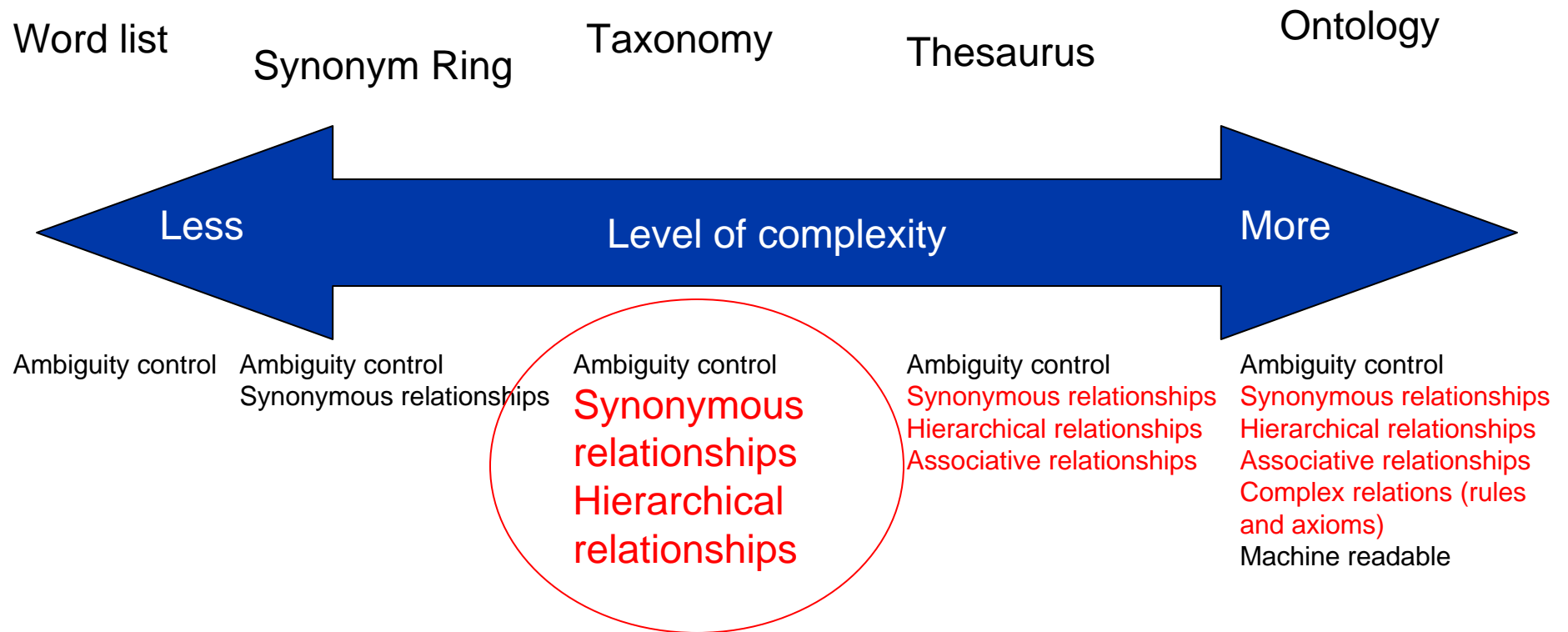
Romanization

- Look at reference sources to find commonly accepted spellings for terms or proper nouns from non-Roman script languages
- Use a romanization system that uses few or no diacritical marks



Relationships

Types of Semantic Relationships



Based on ANSI
Z39.19-2005

Relationships in a Taxonomy

- Broader terms
- Narrower terms
- Equivalent terms
- Associative relationships are not described

Hierarchical Relationships

- Generic

- fruit

- grapes

- Instance

- grapes

- Concord grapes

- Whole/Part

- Canada

- British Columbia

- Vancouver

Generic Relationships

- Often called an “IsA” relationship
 - Fill in the blanks
 - [narrower term] *is a* [broader term]
- Meets requirements of “all-and-some” test
 - All robins are birds; Some birds are robins.
- Some generic relationships just meet the “some-and-some” test
 - Some robins are NW birds. Some NW birds are robins.

Instance Relationship

- Makes link between a general category and an individual instance
- Can also be an “IsA” relationship
 - A Concord grape (NT) is a grape (BT)

Whole/Part Relationship

- One concept is inherently included in another
- Logical hierarchy
- The whole is the broader term
 - Canada
 - British Columbia
 - Vancouver

Exercise

Take controlled list and make it into a taxonomy with broader and narrower terms

Equivalence Relationships

- A concept that can be expressed by two or more terms
- One becomes the preferred term
- The others are the non-preferred terms
- Five types of equivalence relationships

Equivalence Relationships

- Synonyms
 - Different linguistic origins
 - Popular vs. scientific names
 - Generic vs. trade names
 - Emergent concepts
 - Euphemisms
 - Slang and jargon
 - Dialectical variants

Equivalence Relationships

- Near-synonyms
 - Equivalent for the purposes of the depth of the taxonomy

Equivalence Relationships

- Lexical variants
 - Direct vs. inverted order
 - Spelling variants
 - Stem variant
 - Irregular plurals
 - Acronyms

Equivalence Relationships

- Generic posting
- Cross-references to compound term elements

Polyhierarchies

- Concepts fit in more than one category
- Look at generic relationships
- Look at whole-part relationships
- Look at multiple types of relationships

Polyhierarchy Example

fruits

vegetables

Vine fruits

salad vegetables

tomatoes

```
graph TD; fruits --> VF[Vine fruits]; VF --> tomatoes; vegetables --> SV[salad vegetables]; SV --> tomatoes;
```

Scope Notes

- Useful for
 - Explaining the meaning of a term
 - Distinguishing between terms with overlapping meanings
 - Providing advice on term usage
 - air defense has this scope note:
Measures to nullify or reduce effectiveness of attack by airborne enemy airplanes or missiles

History Notes

- Useful for
 - Tracking the usage and meaning of a term over time
 - air cushion vehicles has this history note:
 - For entries prior to 2000, see Ground effect machines

Term Records

- Individual records for each term
 - Term
 - Source
 - Scope note
 - Used for reference
 - Relationships
 - Category or class number
 - History note (date added, modified)
 - Who added/modified

Part 3

Taxonomy Development

Taxonomy Considerations

- Avoid duplicating a vocabulary that already exists
- Consider construction methods
 - Committee
 - Empirical
 - Combination
 - Machine assistance

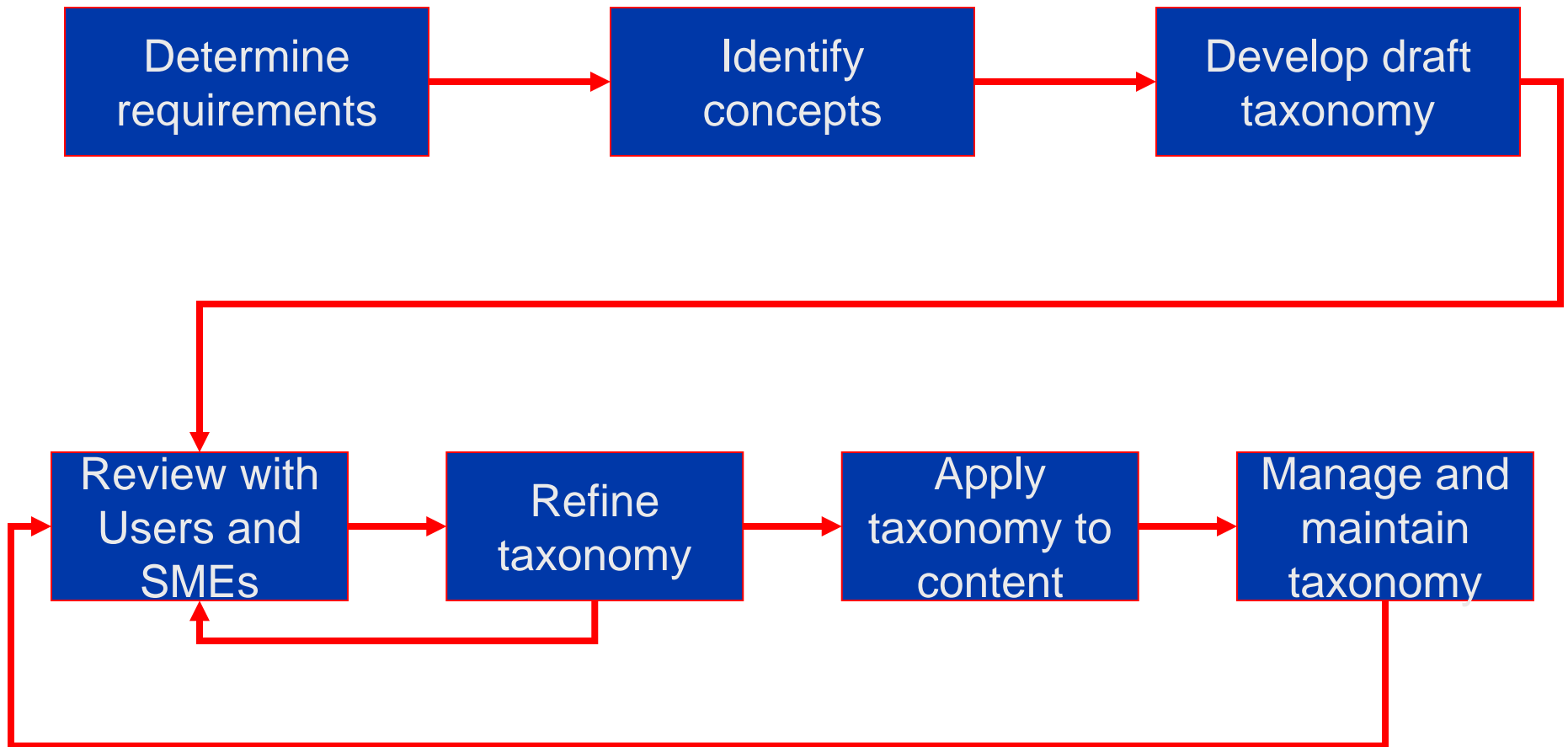
Dimensions of a Taxonomy

- Industry perspective
- Business processes
- Infrastructure

Size of Taxonomies

- Size is dependent upon the user requirements
 - Large taxonomies +20,000 terms
 - May have ten levels or more
 - Subsets of large taxonomies for unique requirements
 - A navigation taxonomy
 - Should not present more than 10 topics
 - Should not have more than 3-4 levels

Building a Taxonomy



Determine Requirements

- Scope
- Purpose
- Format of content
- Subject or facet coverage
- Depth
- Type of content
- Volume of content
- Target audience
- User needs
- Any technology requirements

Identify Concepts

- Identify source materials
- Gather concepts
- Analyze search logs
- Inventory content
- Analyze content
- Determine content types
- Interview SMEs
- Identify existing taxonomies
- Extract candidate terms

Develop Draft Taxonomy

- Develop upper levels of structure
- Start broad, not deep
- Work from bottom up and top down
- Reconcile terminology issues
- Use concepts universally
- Establish common rules

Exercise

- Build Taxonomy with supplied concepts
 - Modify concepts to ANSI standard format if necessary
 - Structure them into a hierarchy
 - Create labels for top nodes, if needed
- Add new terms (will be handed out)

Review with Users and SMEs

- Provide draft for review
- Involve stakeholders, SMEs and users
- Build consensus
- Incorporate responses
- Keep a history of decisions
- Conduct usability studies

Refine Taxonomy

- Incorporate refinements
- Continue review and refine cycle
- Know when to quit
 - Low level of detail vs. value at the leaf node
 - Don't overbuild
- Establish test criteria

Apply Taxonomy to Content

- Provide guidelines for use
- Deploy
 - Navigate web sites
 - Tag content
 - Integrate with existing applications

Manage and Maintain Taxonomy

- Establish ownership
- Establish governance processes
- Create change control processes
- Develop maintenance plan
- Review content for new concepts
- Develop user feedback process for new concepts
- Maintain lifecycle (version control)
- Review success criteria
- Provide documentation

Part 4

Maintenance and Metrics

Why Taxonomies are Constantly Evolving...

- Language changes over time...
- New concepts emerge...
- Terminology and usage change...
- Some terms go out of fashion...

Keep the Taxonomy Up-to-Date

- Review taxonomy periodically for currency
- Create a “candidate” list of terms for consideration
- Analyze items returned in error
- Sample newly added content

Updating the Taxonomy

- Add new terms
- Modify existing terms
- Delete terms
 - Consider terms used excessively or infrequently
- Consider impact of modified or deleted terms
- Coordinate updates with other information systems using the vocabulary
- Monitor any hyperlinks

Taxonomy Change Control

- Provides taxonomy stability
- Ensures taxonomy currency
- Minimizes impact of company reorganization

Types of Automated Tools

- Taxonomy development and maintenance
 - Assist humans in taxonomy creation
 - Data Harmony, TermTree, SchemaLogic
- Classification and categorization
 - Examines sets of documents and tags for categories
 - Autonomy/Verity
- Content companies
 - Combines purchased content with internal content
 - Factiva, LexisNexis

Automated Classification

Factors to Consider

- Diversity of topics covered
- Length and style of documents in the corpus
- Presence or not of controlled vocabulary
- Set of example documents
- Size of the problem
- Resources available
- Ability to incorporate end user feedback
- Business value
- Risk of misclassification
- Payback when balancing automation vs. human intervention

Taxonomy Management System

Key Features

- Non-proprietary
- Hardware independent, standard operating system
- Supports basic term relationships
- Supports scope and history notes
- Allows definition of additional fields in records
- Supports unique IDs
- Complies with filing and sorting guidelines

Taxonomy Management System

Key Features

- Requires little user training
- Allows multiple users
- Displays and prints in both upper case and lower case
- Provides hierarchical and alphabetical displays
- Provides reporting capabilities
- Supports multiple vocabularies
- Provides workflow capabilities
- Provides governance voting

Taxonomy Management System

Term Record Displays

- Term records comply with standards
- No limit on the number of characters in a term
- Distinguishes between preferred and entry terms
- No limit on the number or types of relationships
- Ability to edit a term's position in a hierarchy
- Ability to display poly-hierarchy

Taxonomy Management System

Cross References

- Checks the validity of references
 - No cross-references to dead end terms
 - Non-preferred terms can only have a USE reference
 - A term is not related to itself
 - No conflicting references
 - Eliminates duplicate cross-references
- Creates reciprocal references
- Maintains references reciprocally
 - References to a modified or deleted term are changed automatically

Taxonomy Management System

Term Maintenance

- Prompts for verification when a term is deleted
- Makes reciprocal deletions and prompts if an orphan term is created
- Provides for designation of candidate terms
- Allows for the comparison and merge of terms
- Logs terms used in search queries
- Provides error checking

Taxonomy Management System

Searching

- Allows full term search or truncated search
- Searches for non-preferred and refers to preferred term
- Search not case sensitive
- Allows browse in hierarchical and alphabetical view
- Allows display of complete term record and relationships

Taxonomy Management System Reports

- Term listings
- Term relationships
- Statistical reports
- Customizable reports

Testing the Taxonomy

- Does the taxonomy provide appropriate search results
- Does the taxonomy match user expectations
- Evaluation criteria
 - Should support Taxonomy purpose
- Testing methods
 - Heuristic evaluation (experts evaluate)
 - Affinity modeling (card sorting)
 - Usability testing (overall system)

Factors Affecting Metrics

- Frequent updates to the taxonomy
- Various user needs
- Quality of source material
- Growth of new source material
- The number of sources
- Similarity of sources
- How well the source materials cover the domain

Qualitative Methods for Testing

- Demonstrate to SMEs
- User satisfaction survey
- Usability studies
- Analyze items returned in error
- Tag sample content
- Relevancy testing
- Item reuse

Quantitative Methods for Testing

- How evenly does the taxonomy divide the content?
- Is the indexing repeatable?
- How well does the taxonomy match the content?

Taxonomy Coverage Metrics

- Create list of key concepts from source materials
- Compare with the taxonomy to be evaluated
- Compare with a controlled vocabulary developed for the field

Assessing Node Levels

- Depth of branches
- Average number of children for each node
- Balance indicators between nodes

Assessing Terminology

- Objectivity
 - Term must describe an actual concept
- Duplication of trees
- Consistency of hierarchical relationships

Additional Measures of Taxonomy Quality

- Arrangement of classes
- Cross references within the taxonomy
- Stability of the taxonomy

Precision and Recall

$$\text{Precision} = \frac{N}{X} \qquad \text{Recall} = \frac{N}{Y}$$

- Recall and precision are inversely related
- As recall goes up, precision decreases
- Higher recall – user wants to ensure no critical documents are missed
- Higher precision – user wants to find specific documents

N = number of relevant document retrieved

X = total number of documents retrieved

Y = all existing documents

Precision and Recall

Precision becomes increasingly important as growth in the amount of content to be searched results in increased retrieval of relevant documents

Optimizing Precision and Recall

- Term relationships
- Number of instances per term
- Specificity of terms
- Coordination
- Normalization of terms
- Uniformity and coverage

Taxonomy Evaluation

- Evaluate taxonomy's match to its purpose
- Check usability of electronic versions of taxonomies
- Check sources used during development
- Review procedures used in development
- Examine structure
- Analyze for internal consistency
- Check against authoritative sources
- Examine the taxonomy when applied to search and to indexing



Engagement and Governance

Engaging People

- Who to include in taxonomy development
- Multi-disciplinary team
 - Terminology decisions – Information professionals (librarians, archivists)
 - Technology decisions – IT
 - Subject matter experts from across the company
 - End users
- IT primarily tasked with maintenance of the software
- Taxonomy is the responsibility of the SMEs, content owners and librarians. (Delphi 2004)

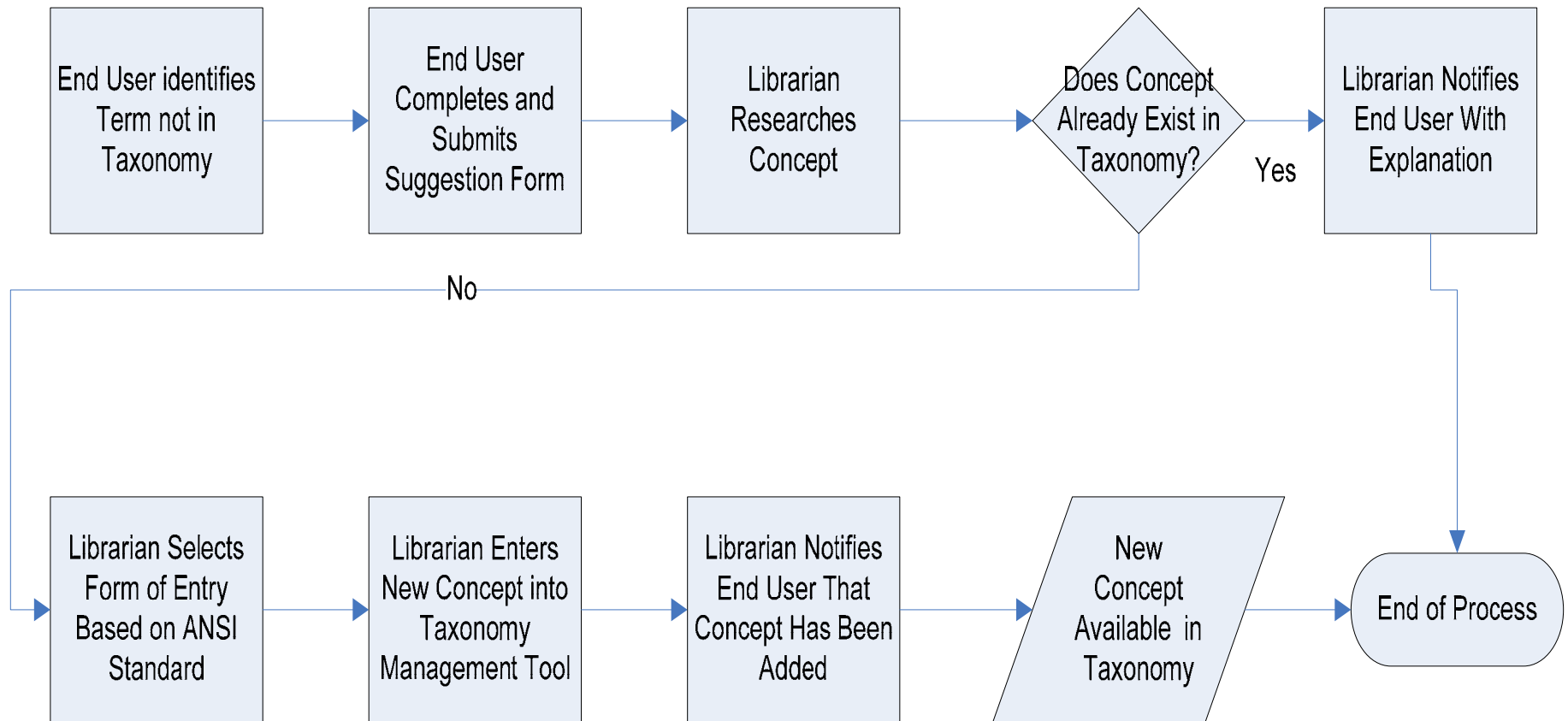
Buy-In with End Users

- Come to an agreement
- Negotiate as needed
- Include a process for feedback

Exercise

Collaborate on taxonomy from previous exercise

Suggest a Taxonomy Term Process Flow



Suggest a Term Form

Please fill in the form below as completely as possible. Include one Broader Term (a "parent" concept). This establishes a relationship between the terms, either generic (class and species), whole/part, or an "is a" instance relationship.

Term	<input type="text"/>
Definition	<input type="text"/>
Broader Term(s)	<input type="text"/>
Narrower Term(s)	<input type="text"/>
Synonym(s)	<input type="text"/>
Notes	<input type="text"/>
Source Document or URL	<input type="text"/>
Your name	<input type="text"/>
Your email	<input type="text"/>
	<input type="button" value="Submit the Term"/> <input type="button" value="Reset"/>

Taxonomy Governance Processes

- Taxonomy team reports to a Change Board
- Implement a comment handling process
- Implement an appeals process
- Use issue logs
- Develop a communication plan
 - Announcements of enhancements, releases, Board activities, etc.
- Develop a release schedule

Governance Documentation

- Charters
- Processes
- Goals
- Metrics
- Communication plans
- Presentations
- Training materials

Sample Taxonomy Team Charter

- Purpose
 - Develop and maintain taxonomies
- Scope
 - Review, approve, and implement changes
- Deliverables
 - Process documentation
 - Documentation of decisions and meeting minutes
 - Documentation regarding authoritative standards used, guidelines and best practices
 - User interface to the taxonomy
 - The taxonomy itself!!

Sample Taxonomy Team Charter page 2

■ Responsibilities

- Review and manage requests for changes to the Taxonomy
- Maintain a change log to track requests for changes
- Work with subject matter experts
- Communicate implemented changes
- Educate users and projects about the Taxonomy
- Develop and maintain the Taxonomy web site
- Evaluate and recommend back-end taxonomy management tools
- Attend meetings prepared to discuss items on the agenda
- Complete assigned responsibilities in a timely manner
- Promote Taxonomy activities in the enterprise

Sample Taxonomy Team Charter page 3

- Membership
 - Members must have expertise in vocabulary management
- Authority
 - Team receives its authority from Library Management
- Accountability
 - Team is accountable to Library Management and the Taxonomy Change Board
- Voting
 - Team strives for understanding, acceptance and support
 - Decisions reached by consensus whenever possible
 - When voting, each member has one vote and all will abide by the outcome. A quorum is required for voting.

Sample Change Board Charter

- Purpose

- Reviews and resolves issues related to the Taxonomy that are not resolved at the Taxonomy Team level. Reviews and approves major changes to the Taxonomy

- Scope

- Reviews and approves major changes to content and structure of the Taxonomy. Reviews and approves changes to strategy and functionality. Elevates issues to Advisory Board for guidance.

- Deliverables

- Process documentation for Change Board activities
- Documentation for Taxonomy strategy and approach
- Documentation of Change Board decisions and meeting minutes

Sample Change Board Charter page 2

■ Responsibilities

- Address and resolve issues on which the Taxonomy Team is unable to come to consensus
- Review and approve major content changes (for example new facets and high level structural changes)
- Communicate implemented changes to the Advisory Board
- Interface with Advisory Board members about the Taxonomy
- Attend meetings prepared to discuss items on the agenda
- Complete assigned responsibilities in a timely manner
- Promote Taxonomy activities in the enterprise
- Educate users and potential customers about the Taxonomy

■ Membership

- Voting members must have expertise in vocabulary management
- Representatives from business functions (non-voting)
- Taxonomy Team representatives (3 members)

■ Management

Sample Change Board Charter page 3

- Authority
 - Change Board receives its authority from management
- Accountability
 - Accountable to management
- Voting
 - Change Board strives for understanding, acceptance and support
 - Decisions reached by consensus whenever possible
 - When voting, each member has one vote and all will abide by the outcome. A quorum is required for voting.

Sample Advisory Board Charter

- Purpose
 - Provide strategic direction and promote the taxonomy
- Scope
 - Provide guidance on strategy and functionality
 - Serve as a conduit between business and the team
- Deliverables
 - Documentation of Advisory Board minutes
- Responsibilities
 - Provide future strategic direction
 - Promote and support taxonomy in the enterprise
 - Act as a conduit between potential customers, users, and the taxonomy team

Sample Advisory Board Charter page 2

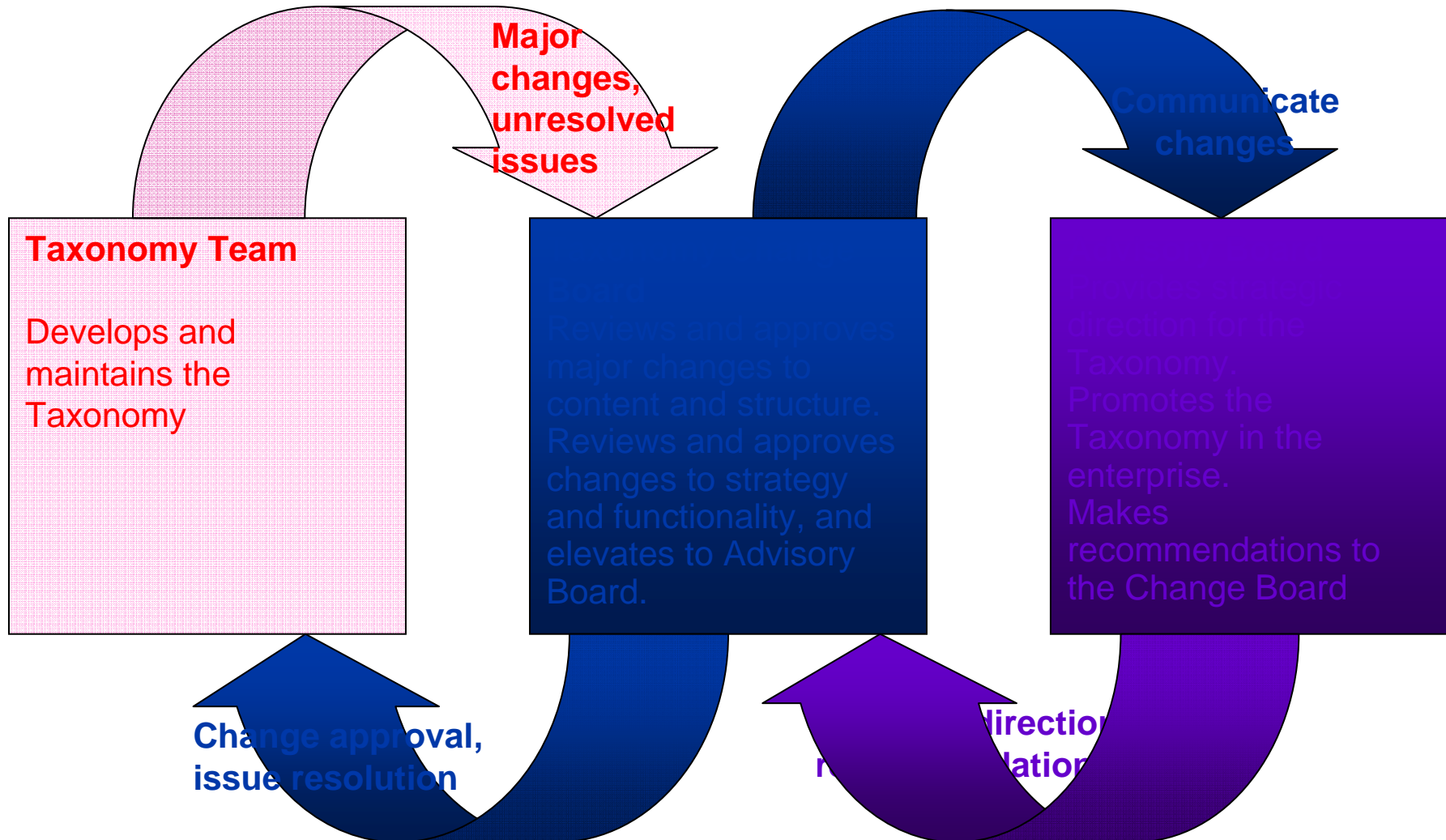
■ Membership

- Voting members must understand use and value of controlled vocabularies.
- Should be opinion leaders in information management
- Able to clearly communicate and promote the Taxonomy
- Should be five members with representatives from the business units and one Library representative. Selection based on Advisory Board recommendations
- Meet quarterly

■ Authority

- The Advisory Board receives its authority from the Libraries

Governance Flow



Taxonomy Process Documentation

- Editorial rules and guidelines
- Training materials
- Processes and workflows
- Marketing materials
- Presentation materials
- Conflict resolution process
- Taxonomy metrics
- ROI analysis



Applications and Interoperability

Facets

- A dimension or feature type
- Useful for
 - Emerging fields
 - Interdisciplinary areas
 - Vocabularies of multiple hierarchies
 - Classifying electronic content

Development of Facets

- Can often use a phased approach
- Factors to consider
 - How and how soon will the facet be used
 - How helpful will it be to users
 - How easy will it be to develop the vocabulary
 - Does the team have expertise
 - Are there metadata tagging requirements

Metadata Definition

- Characteristics of a content item
- Attributes and properties about the item
- Metadata tags are applied to describe a content item

Metadata Schemas

- Metadata element sets specify semantic rules of each element
- Metadata schemas add rules for syntax (form and structure)
- Facilitates interoperability
 - Rely on controlled vocabularies for element content

Examples of Metadata Schemas

- Dublin Core
 - Content (what it is about)
 - Context (administration information)
 - Structure (relationships with other content)
- IMS Learning Resource Meta-data
- MARC Standards

Sample Controlled Vocabularies

Object Facets	Definition	Possible Sources
Organization	Names of organization	Name authority; SP 800-87; Organization structure
Location	Names of geographical areas	ISO 3166 Country names; Postal codes
Industry	Broad business categories	NAICS
Format	Physical manifestation	MIME type, Internet Assigned Numbers Authority
Language	Language of the object	RFC3066; ISO 639
Content Type	Types of content being managed	Dublin Core Type Vocabulary; AGLS Document Types

Metadata Tagging & Controlled Vocabularies

Creator	Kathryn Breininger	Supplied by system
Title	Using metadata to organize information	Free form entry
Date	04/11/2006	Supplied by system
Subjects	Metadata, Tagging	From subject taxonomy
Category	Information management	From process taxonomy
Geographic location	Seattle, WA	From geographic taxonomy
Project	Content Management Initiative	From project taxonomy
Organization	Library Services	From name vocabulary

Searching with Controlled Vocabularies

Creator

breininger

Pick from name authority

Title

Free form entry

Project

Taxonomy
Initiative

Pick from project
taxonomy

Subjects

metatagging

Pick from subject
taxonomy

Controlled Vocabularies and Interoperability

- Communities develop controlled vocabularies specific to their domain and terminology
- Searchers want a single search in databases serving different domains
- Interoperability
 - Relate controlled vocabularies that were intended to stand alone
 - Permits searches from one vocabulary to be effective in another

Interoperability and Compatibility

- Searching compatibility
- Merging databases
- Merging controlled vocabularies

Achieving Interoperability

- Modeling
- Translation
- Satellite
- Node linking
- Mapping
- Co-occurrence mapping
- Switching
- Temporary linking
- Linking controlled vocabulary servers


From ANSI Z39.19-2005 Appendix D

Maintenance in Multiple Controlled Vocabularies

- Authority records
- Vocabulary mapping
- Semantic network
- Lexical database

Factors Affecting Interoperability

- Similarity of content subject matter
- Precoordinated vs postcoordinated
- Level of specificity
- Synonym treatment
- Search methodologies
- Warrant taken
- Intended purpose



Strategy and Implementation

Taxonomy Implementation Drivers

- Inadequacy of search
- Inability to find information has impact on a business' bottom line
- IT has “discovered” taxonomy from vendors
- Support for Content Management Systems
- Support of Knowledge Management initiatives

Taxonomy Strategy


- Integrated with business strategy
- Integrate with search functions
- Leverage strategically
- Integrate multiple corporate taxonomies
- Strategy for taxonomy should reside with those who do classification

Additional Considerations

- Keyword searching in Taxonomy
- Ability to view full term records
- Pick lists or drop down menus
- Web navigation
 - Use hyperlinks to navigate inside the taxonomy
 - Provide means to browse the taxonomy

Features of a Successful Taxonomy

- Extensible over time
- Uses a taxonomy management system independent of applications
- Connected to content creators
- Takes into account business priorities



Benefits and Best Practices

Productivity Benefits

- More time for analysis
- Less duplication of effort
- Fewer poor decisions
- Information is a strategic asset
- More internal knowledge sharing
- Common understanding of terminology

Searching Benefits

- Creates structure in unstructured content
- Facilitates navigation among terms
- Provides powerful browse capabilities
- Provides standardized access
- Organizes fragmented content

Business Benefits

- Reflects scope of your business
- Assists learning in the domain
- Supports business goals

Taxonomy Value

40% of respondents to a Delphi survey identified the primary value of taxonomies as being the navigation tool for the discovery of content.

Taxonomies and ROI Considerations

- Building and maintaining a taxonomy is a cost
- Tagged content must be available and used
- Technology costs associated with building a taxonomy
- Integrate taxonomies with applications for maximum return on investment

Taxonomy ROI Examples

- Increased sales through improved access to online product catalog
- Improved efficiencies in Help or Call Centers
- Knowledge worker productivity

Taxonomy Development Best Practices

- Find a strong sponsor and champion
- Establish a team to manage taxonomies
- Start with an overall framework
- Use automatic approaches to facilitate manual development – need human review
- Develop taxonomies for multiple facets
- Establish quality control measures and metrics
- Implement a phased rollout

Taxonomy Development

Things to Avoid

- Do not use corporate lingo
- Do not use departmental names
- Do not use undefined acronyms
- Avoid over thinking
- Avoid over-engineering
- Avoid developing unneeded sections
- Avoid using a proprietary system for taxonomy management

Taxonomies in the Information Lifecycle

- Implement taxonomies into the information life cycle
- Use taxonomies to tag content at information creation

Critical Success Factors

- Understand target audience, requirements
- Develop strong stakeholder and user relationships
- Assign responsibility for ownership and maintenance
- Research lessons learned from other companies

Critical Success Factors

- Understand how tools and technology play
- Build, buy and re-use
- Build in flexibility
- Differentiate between the taxonomy and the metadata scheme
- Develop metrics on value and relevance of the taxonomy
- Plan for long term – have a vision and strategy
- Plan for maintenance resources

Questions?

Thank you!

Contact Information

- Kathryn Breininger

kathryn.r.breininger@boeing.com

- Mary Whittaker

mary.s.whittaker@boeing.com