

# Controlled Vocabularies: Where to start when you don't know what to do

Presented by John Magee,  
Director of Indexing & Vocabulary Services,  
Cengage | Gale  
Council of Science Editors Conference  
May 6, 2019

EMPOWER™ CHANGE



## Controlled Vocabularies: Where to start when you don't know what to do

Usually, the starting line on the path towards controlled metadata is a comment like this, "I think we need more organization here."

But what sort of controlled metadata do you need?

**There are many forms of controlled descriptive metadata:**

- Controlled list
- Look-up table
- Synonym ring
- Lexicon
- Folksonomy
- Taxonomy
- Authority control file
- Faceted concepts
- Thesaurus
- Ontology

EMPOWER™ CHANGE



## Controlled Vocabularies: Where to start when you don't know what to do

### HOW WILL PEOPLE USE THIS VOCABULARY?

- Where to start? Start with your users' needs.

### • HOW WILL PEOPLE USE THIS VOCABULARY?

- Ask questions. Lots of questions.
- But always bring it back to your users.

## Controlled Vocabularies: Where to start when you don't know what to do

### HOW WILL PEOPLE USE THIS VOCABULARY?

“I need a controlled vocabulary so that I can \_\_\_\_\_.”

- Organize a product catalog
- Provide a search filter
- Send content to different areas
- Provide personalization for users
- Supplement search
- Provide synonyms and alternative names
- Link similar pages
- Provide a basis for custom queries

## Controlled Vocabularies: Where to start when you don't know what to do

### THINGS TO CONSIDER

- Hierarchy
- Granularity
- User syntax and language
- Ongoing maintenance
- Application of terms to content
- Required mappings to other vocabularies
- Budget considerations

## Controlled Vocabularies: Where to start when you don't know what to do

### THINGS TO CONSIDER: HIERARCHY

- Levels/Depth of relationships
- Simple or complex hierarchy?
- Polyhierarchy?

#### Questions:

- How will our users navigate this vocabulary?
- Are there specific UI features that will rely on the hierarchy?
- Will this structure be used to roll up content in some places?
- Will that happen in ways that our users expect?

## Controlled Vocabularies: Where to start when you don't know what to do

### THINGS TO CONSIDER: GRANULARITY

- Total number of terms
- Specificity of concepts

#### Questions:

- How many items are we organizing?
- How many items will users expect to see with each term or concept?
- What is an acceptable retrieval-set size?
- How important is it to distinguish among similar but discrete concepts?
  - This is an especially user-centric point. 13-inch beagles and 15-inch beagles are all just beagles to most folks. But that difference is very important at the Westminster Dog Show.

## Controlled Vocabularies: Where to start when you don't know what to do

### THINGS TO CONSIDER: USER SYNTAX AND LANGUAGE

- Reading level
- Technical terms and jargon
  - "Dihydrogen monoxide (DHMO)" UF "Hydrogen hydroxide"
- Language / Multilingual
- Form of entry
  - Noun v. Verb – "Oscillations" vs. "Oscillating"
  - Plural/singular – "Wave" v. "Waves"
  - Inverted formats – "Isaac Newton" vs. "Newton, Isaac"
  - Honorifics – "Newton, Isaac, Sir" vs. "Newton, Isaac"
  - Modifiers – "Isaac Newton" vs. "Isaac Newton (English Physicist)" vs. "Isaac Newton (b. 1642 - d. 1727)" vs. "Isaac Newton (physicist, 1642-1727)"
  - Nested context or hierarchy ("Mammals/Dogs/Beagles/13-Inch" vs. "Mammals" NT "Dogs" NT "Beagles" NT "13-Inch Beagles")

#### Questions

- What will our users see? Will this vocabulary be visible to them?
- What language will my users expect? What syntax?

## Controlled Vocabularies: Where to start when you don't know what to do

### THINGS TO CONSIDER: ONGOING MAINTENANCE

- Static list?
- Growing/evolving vocabulary?

**The simpler the vocabulary, the easier it is to maintain over time**

#### Questions

- What is the long-term plan for this vocabulary?
- Will it grow and change with our audience?
- Who will maintain this vocabulary?
- Where will this vocabulary live? (Offline spreadsheet? Vocabulary management system? Lookup table within a block of coding?)

## Controlled Vocabularies: Where to start when you don't know what to do

### THINGS TO CONSIDER: APPLICATION OF TERMS TO CONTENT

- Content creation process
- Post-content tagging
- Manual tagging
- Automated tagging

#### Questions:

- How will my end users access and use this content?
- Will somebody be happy or sad if they see this item with this term?
- How am I going to apply this vocabulary to my content?
- Who does this work? What tools do they use?
- And remember: these folks are also users of this vocabulary.

## Controlled Vocabularies: Where to start when you don't know what to do

### THINGS TO CONSIDER: REQUIRED MAPPINGS TO OTHER VOCABULARIES

- External standards
- Library of Congress, Wikidata, VIAF, Getty AAT, MeSH, etc.
- Internal vocabularies
- Legacy vocabularies
- Additional uses (catalogs, business systems, other products)

#### Questions

- Will users expect to cross-search?
- Has our content already been tagged with another vocabulary?
- Will users expect to use that vocabulary, too?

## Controlled Vocabularies: Where to start when you don't know what to do

### THINGS TO CONSIDER: BUDGET AND TIME CONSIDERATIONS

I put this last on purpose. Start by taking this out of consideration. It's an enormous and genuine constraint. But start by defining the best vocabulary for the requirement before you start trimming corners. That way you'll know where you can trim with the least impact on your users.

#### Questions

- How many things do you have to organize and/or tag?
- When do you need it done?
- What is the most cost-effective way to achieve our goal?
- Is there an existing vocabulary we can use or adapt for this purpose?

**Controlled Vocabularies: Where to start when you don't know what to do**

**ABOVE ALL ELSE, REMEMBER TO ASK THE MOST IMPORTANT QUESTION...**

**How will people  
use this vocabulary?**

EMPOWER™ CHANGE

13

 **GALE**  
A Cengage Company