

You Say Dog, I Say Canine:
Designing a Thesaurus to Connect Your Users to Your Digital Content

Presented NDLA 2010 Annual Conference, September 29, 2010
Aubrey Madler, MLS, Rural Assistance Center &
Maren Niemeier, MLIS, Health Workforce Information Center

Steps to Create a Controlled Vocabulary (CV)

1. Generate Wordstock
Include subjects, names (place names, personal names, organization names), format (physical material, size, color), time (date, season)
2. Decide on Format, Choose Preferred Terms & Identify Synonyms
Resolve synonyms (Use/UF – use for), capitalization, abbreviations, plurality, abstract concepts, compound terms
3. Choose Hierarchies and Facets
Identify hierarchies and list narrower terms (NT), broader terms (BT)
Select facets, which are homogenous or mutually exclusive categories of terms
4. Add Associative Relationships
Connect concepts using “See also” and related term (RT)
5. Select Thesaurus Design and Display
Electronic/print, external/internal view, facets, BT/NT/RT, topical/alphabetical, visual/textual

Tips to Jump Start Your Project

Your resources

Why do you want a controlled vocabulary? Is a CV necessary? Are there alternative solutions?
Do you have the resources to develop and maintain a CV? Training, time, staff, software...
Are you prepared for future growth/additional terms?

Your content

What are your topics?
Will the collection’s coverage change over time?
What sources can you use for wordstock?
What is the broadest level of vocabulary you could use for this project?

Your users

Who are the potential users of this collection?
Do you need different entry terms for different audiences?
Who will be indexing materials and using the controlled vocabulary?

Further Reading:

ANSI/NISO Z39.19-2005: *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, American National Standards Institute

http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a

-guidelines and conventions for the contents, display, construction, testing, maintenance, and management of monolingual controlled vocabularies.

Cataloging for the 21st Century—Course 3: Controlled Vocabulary and Thesaurus Design

<http://www.loc.gov/catworkshop/courses/thesaurus/index.html>

-course content from the Library of Congress that addresses many aspects of CV and thesaurus design

Controlled Vocabulary: Example Controlled Vocabularies and Thesauri

<http://www.controlledvocabulary.com/examples.html>

-further defines controlled vocabularies and offers links to various CV examples

Introductory Tutorial on Thesaurus Construction

<http://publish.uwo.ca/~craven/677/thesaur/main00.htm#contents>

-module available from the University of Western Ontario reiterates points from our presentation and offers additional examples and explanations

Future Directions and Additional Concepts: You have your terms, now what?

Descriptions of Metadata Schema and Digital Library Standards: <http://www.loc.gov/standards/>

RDF and Semantic Web: *Quick Intro to RDF* <http://www.rdfabout.com/quickintro.xpd>, “*What is RDF and what is it good for?*” <http://www.rdfabout.com/intro/?section=contents>, “*Semantic Web FAQ (W3C)*”: <http://www.w3.org/RDF/FAQ>

Linked Data: Wikipedia article: http://en.wikipedia.org/wiki/Linked_Data, Linked Data Web site: <http://linkeddata.org/>

Software and Content Management System (CMS) Options:

Comparison of CMS Software: <http://cms-software-review.toptenreviews.com/>, <http://php.opensourcecms.com/>,

Drupal (open source CMS): <http://drupal.org/>