



Integrating, Merging and Mapping Controlled Vocabularies

Heather Hedden

Hedden Information Management

www.hedden-information.com

Integrating, Merging or Mapping

- **Integrating:**
Combining separate controlled vocabularies (CVs) into a single, larger master controlled CV for combined use



Integrating, Merging or Mapping

■ Merging:

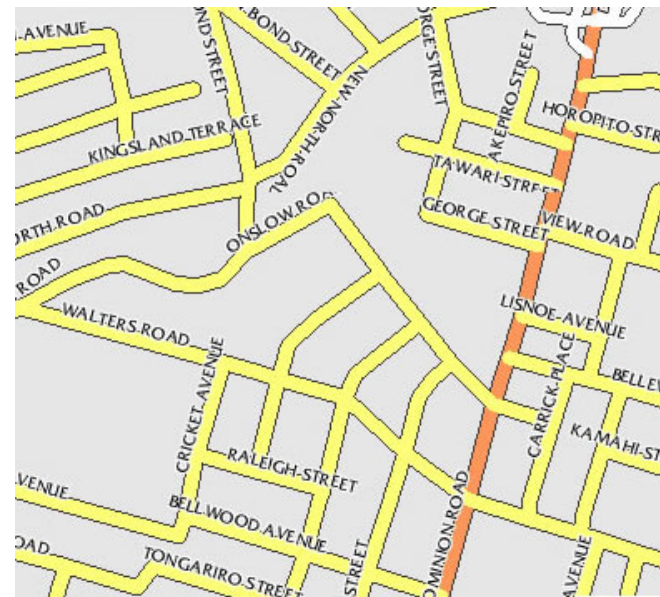
Combining two or more redundant vocabularies in same subject area into one

- Without any longer retaining them as distinct
- Legacy content is retrieved through added equivalence relationships



Integrating, Merging or Mapping

- **Mapping:**
 - Enabling one CV to be used for another in same subject area
 - Retain them both as continued distinct vocabularies.
 - A CV continues to be used to retrieve its content as before plus additional content associated with the other.



Something representing something else

Integrating Vocabularies



- Combines related, but not redundant taxonomies
- Increases (multiplies) the number of preferred terms
- Involves importing an additional taxonomy/hierarchy/thesaurus into a taxonomy management system to be used with other taxonomies
- Concerned with issues of interoperability
- Taxonomist deals with integrating structures more than integrating individual terms.

Integrating – Situations



- An enterprise taxonomy is built via combining existing departmental taxonomies.
- An additional facet is added to an existing faceted taxonomy.
- A new product line is added, requiring a new product/topic hierarchy.
- An organization acquires/merges with another organization, and their CVs in different areas are integrated.
- An internally created CV is supplemented by a purchased/licensed CV in a complementary subject area to expand its scope.



Merging and Mapping Vocabularies

- Compares two closely redundant vocabularies side-by-side, term-by-term
- First pass(es) automatic followed by taxonomist review of matches
- Taxonomy software may have the feature (Synaptica, Wordmap), or do your own scripting
- Taxonomist reviews, discerns distinction between equivalent, broader/narrower, related terms to approve matches
- Taxonomist deals with terms more than structure.

Merging – Which Direction?



Designate a dominant/primary CV into which to merge the other:

- If an organization acquires another, then the acquirer's CV is dominant.

Or choose:


- The larger CV
- The CV with greater breadth
- The CV with greater depth
- The more structured CV
- The “better” CV



Merging – Automated matching



Use a software tool to compare vocabularies,
to obtain matches in succeeding passes:

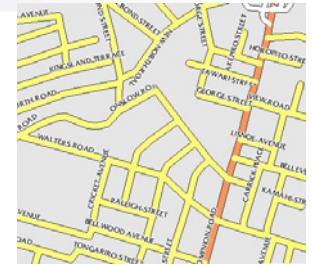
Primary CV	Merging CV	Taxonomist Reviews
Exact matches of:		
<i>Preferred term</i>	<i>Preferred term</i>	
<i>Nonpreferred term</i>	<i>Preferred term</i>	
<i>Nonpreferred term</i>	<i>Preferred term</i>	yes
<i>Nonpreferred term</i>	<i>Nonpreferred term</i>	yes
Inexact matches of:		
<i>Preferred term</i>	<i>Preferred term</i>	yes

Merging – Situations



- An enterprise taxonomy replaces multiple CVs of separate administrative departments
- An organization acquires or merges with another organization, and their redundant vocabularies are merged
- A folksonomy is incorporated into a CV
- An internally created CV is combined with a purchase/licensed CV

Mapping



- Between a retrieval/user-interface CV and a CV indexed to content
- Unmatched terms cannot be utilized.
- Narrower-to-broader matches are fine.
- Includes same kinds of matches as in CV merging, plus matches of words/phrases (of the retrieval taxonomy) within a term from the indexing CV

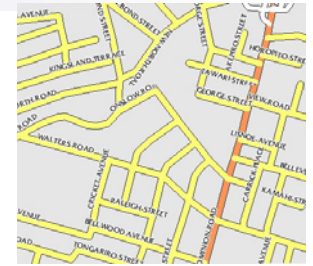


	A	B	C
1	Programmable logic controllers	ok	Programmable controllers
2	Programmable logic devices	ok	PLDs (Programmable logic devices)
3	Programming (Computers)	ok	Computer programming
4	Progressivism (United States politics)	b	Progressive movement
5	Prohibited books	ok	Banned books
6	Project method in teaching	ok	Project method (Education)
7	Projectile points	ok	Projectile points (Archaeology)
8	Projection	n	Projection (Drawing)
9	Projection televisions	ok	Projection television sets
10	Prolactin	n	Prolactin test
11	Proletariat	ok	Working class
12	Prolog (Computer program language)	ok	Prolog (Programming language)
13	Promethazine hydrochloride	b	Promethazine
14	Promoters (Entertainment)	b	Promoters
15	Promotion (School)	ok	Student promotion
16	Pronghorn antelope	ok	Pronghorns
17	Propaganda, American	ok	American propaganda

Indexing CV in column A, retrieval CV in column C,
taxonomist notes in column B

ok is equivalent, b is broader so also ok for upward posting, and n is not acceptable.

Mapping - Situations



- Selected content with an enterprise CV is made available on a public web site with a different public-facing CV
- A content provider with a CV partners with a third-party information vendor with its own CV
- A provider of scientific/technical/medical content with a technical CV creates a simpler CV aimed at laypeople
- Search log query terms need to be integrated into the CV as additional nonpreferred terms.
- To support “federated search” that involves taxonomies

Questions/Contact



Heather Hedden
Hedden Information Management
98 East Riding Dr.
Carlisle, MA 01741
978-371-0822
978-467-5195 (mobile)
Heather@hedden.net
www.hedden-information.com

More Info:

Heather Hedden, *The Accidental Taxonomist*
forthcoming book from Information Today Inc., spring 2010