Managing Mature Taxonomies: Resolving Orphan Terms

SLA Taxonomy Division Webinar
December 12, 2016

Heather Hedden
Senior Vocabulary Editor
Metadata Standards and Services
Gale | Cengage Learning

Heather Hedden

- Senior vocabulary editor, Cengage Learning, 1996-2004, 2014-present
- Author of *The Accidental Taxonomist* (2010, 2016)
- Continuing education instructor, Simmons College School of Library and Information Science
- Former taxonomy consultant

Gale, a Cengage Learning Company

- Subscription databases to libraries: GVRL ebooks, In Context, Academic OneFile, Business Collection, Literature Resource Center, etc.
- Web products to the public: Questia, Books & Authors, HighBeam Research, Encyclopedia.com
- Gale Research reference books, directories, and other book imprints (Greenhaven, Thorndike, St. James Press, etc.)
- Primary Source Media digital archives (Artemis)
- Legacy library database vendor companies: Information Access Company, Predicasts

Academic Journals (347)

Search Terms: *Subject* (controlled vocabularies)

Sort by: Newest

Abstract: *ARTICLE*

Information retrieval in systematic reviews: a case study of the crime prevention literature.

Lisa Tompson and Jyoti Belur.

Journal of Experimental Criminology.

12.2 *(June 2016)* p187. *Word Count: 304.*

Save

Full-text: *ARTICLE*

Reengineering of MeSH thesauri for term selection to optimize literature retrieval and knowledge reconstruction in support of stem cell research.

Yan Su, James Andrews, Hong Huang, Yue Wang, Liangliang Kong, Peter Cannon,, et al.

BMC Medical Informatics and Decision Making.

16.54 *(May 23, 2016) Word Count: 5580.* Reading Level (Lexile): 1440.

Author(s): Yan Su1 , James Andrews2 , Hong Huang2 , Yue Wang1 , Liangliang Kong1 , Peter Cannon2 and Ping Xu1 Background In this "big data" era, the richness of information and the speed at which it spreads is nearly...

PDF page(s) Save

Full-text: *ARTICLE*

Representation of anatomy in online atlases and databases: a survey and collection

Content Types

Academic Journals (347)

Magazines (21)

News (47)

Search within results

Limit Search by:

- Full Text
- Peer Reviewed Journals
- Contains Images

Publication Dates:

Past Week

Past Month

Past Year

Custom Date Range

Subjects

Controlled vocabular... (347)

Genomics (19)

Knowledge management (16)

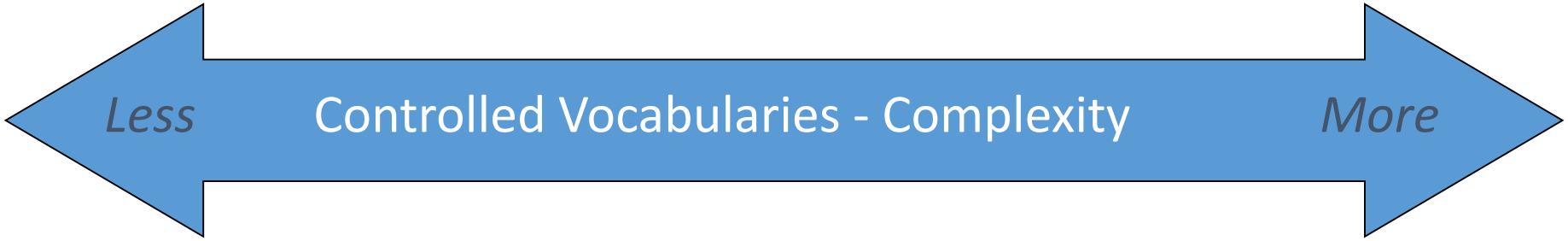
Document Type

Outline

- Taxonomies, Thesauri, and Orphan Terms
- The Gale Project to Review Orphan Terms
- Issues in Finding Parents to Orphan Terms

Taxonomies, Thesauri, and Orphan Terms

Taxonomies and Thesauri Compared



Pick List	Synonym Ring	Authority File	Taxonomy	Thesaurus	Ontology
Ambiguity control	Synonym control	Ambiguity control Synonym control	Ambiguity control (Synonym control) Hierarchical relationships	Ambiguity control Synonym control Hierarchical relationship Associative relationships	Ambiguity control (Synonym control) Semantic relationships Classes

Taxonomies and Thesauri Compared

Taxonomies

- All terms belong to a limited number of major hierarchies (or facets)
- May bend standard hierarchical rules.
- Supports classification, categorization, and concept organization.
  (Like Linnaean taxonomy.)
- Approach is a top-down navigation.

Thesauri

- All terms have relationships, but "hierarchies" can comprise as few as 2 terms.
- ANSI/NISO or ISO standard rules are strictly followed.
- Supports concept scoping, disambiguation, and relationships with similar concepts.
  (Like looking up in Roget's.)
- Approach is term-centered and what terms are linked to/from it.

Taxonomies, Thesauri, and Orphan Terms

Hierarchical Relationship Rules (ANSI/NISO Z.39.19 Guidelines)

1. Generic-Specific

Category or class
NT members/types

Narrower term "is/are a kind of" broader term.

Plants
NT **Trees**

2. Generic-Instance

Common noun
NT Proper noun

Narrower term is an instance of broader term.

Smartphones
NT **Samsung Galaxy**

3. Whole-Part

Concept or entity
NT Part or sub-entity

Narrower term 'is in" broader term (as an integral part).

France
NT **Paris**

Digestive system
NT **Stomach**

Taxonomies, Thesauri, and Orphan Terms

Orphan Term Definitions

1. Terms with no hierarchical or associative relationships (ANSI/NISO Z.39.19 definition)
   - Not permitted in taxonomies or thesauri
2. Terms with no hierarchical (broader or narrower) relationships (“hierarchical orphans”)
   - Not permitted in taxonomies; may be permitted in thesauri
3. Terms with no broader terms (no broader/parent, thus “orphans”) that are not intended as top terms
   - Not desired in taxonomies or thesauri

The problem:
Due to the lack of relationships to other term, orphan terms cannot be found by users when browsing the taxonomy/thesaurus. (Can be found by search, though)

Gale Orphan Term Review Project

Gale Subject Thesaurus

- Used along with multiple separate name authority files and other classification metadata for indexing articles and various other content resources
- 60,000 preferred terms and always growing
- Managed by four vocabulary editors, divided by broad subject area
- Terms belong to one or more of 6 subject areas: Business, Health/Medicine, Humanities, Social Sciences, Science Technology
- Developed in the 1970s based on LCSH
- Thoroughly revised in the early 2000s to become an ANSI/NISO Z.39.19-compliant thesaurus
  - Project changed *See also* relationships to either BT/NT or to RT as appropriate.
  - If terms were left as hierarchical orphans, that was ignored.

Gale Orphan Term Review Project

Orphan Term Review Project Background

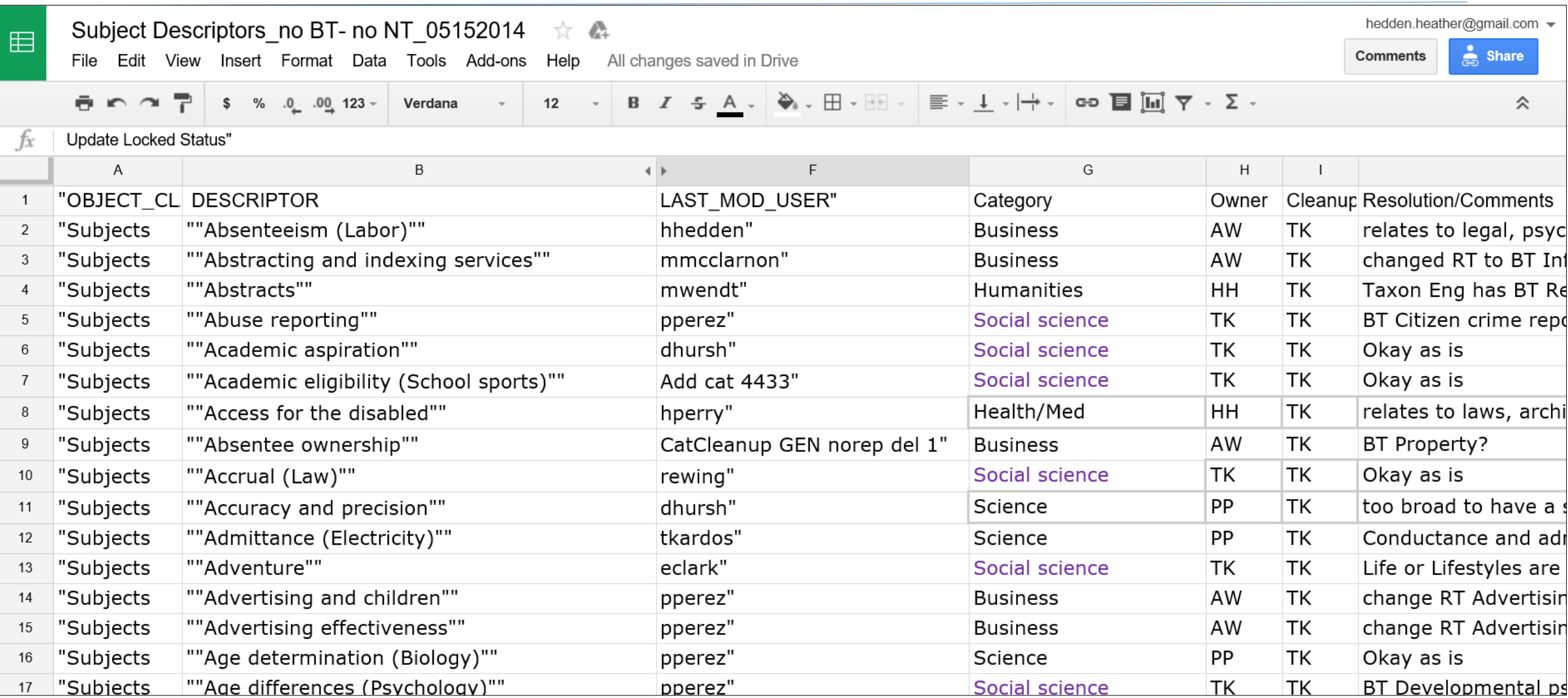
1. Orphan terms with lacking any relationships (hierarchical or associative)
   - Thesaurus management software has report option for this kind of “orphans”
   - Vocabulary editors can/should periodically run reports on their sections of the vocabulary to clean up these kinds of orphans, which are always unacceptable.
2. “Orphan” terms lacking only broader terms
   - A back-end system report needs to be run for this
   - Vocabulary editors review the report to either approve these terms as top terms and/or to add relationships to them.

Gale Orphan Term Review Project

Orphan Term Review Project Background

- Started as a project in April 2014 when a new vocabulary editor joined the team.
- A “back burner” project for vocabulary editors to work on when they are not busy with higher priorities. No timeline or deadline.
- Identified 2420 “orphan” terms (those with no BTs), put them in a spreadsheet
- Two people split the list and provided an initial review with recommendations of broader terms, where applicable, or comments.
- The orphan term list was sorted by subject category and sub-lists of orphan terms for each category were assigned to each vocabulary editor for more detail.

Gale Orphan Term Review Project



Subject Descriptors_no BT- no NT_05152014

hedden.heather@gmail.com

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

$ % .0 .00 123 Verdana 12

fx Update Locked Status"

	A	B	F	G	H	I	
1	"OBJECT_CL	DESCRIPTOR	LAST_MOD_USER"	Category	Owner	Cleanup	Resolution/Comments
2	"Subjects	""Absenteeism (Labor)""	hhedden"	Business	AW	TK	relates to legal, psyc
3	"Subjects	""Abstracting and indexing services""	mmcclarnon"	Business	AW	TK	changed RT to BT Int
4	"Subjects	""Abstracts""	mwendt"	Humanities	HH	TK	Taxon Eng has BT Re
5	"Subjects	""Abuse reporting""	pperez"	Social science	TK	TK	BT Citizen crime repo
6	"Subjects	""Academic aspiration""	dhursh"	Social science	TK	TK	Okay as is
7	"Subjects	""Academic eligibility (School sports)""	Add cat 4433"	Social science	TK	TK	Okay as is
8	"Subjects	""Access for the disabled""	hperry"	Health/Med	HH	TK	relates to laws, archi
9	"Subjects	""Absentee ownership""	CatCleanup GEN norep del 1"	Business	AW	TK	BT Property?
10	"Subjects	""Accrual (Law)""	rewing"	Social science	TK	TK	Okay as is
11	"Subjects	""Accuracy and precision""	dhursh"	Science	PP	TK	too broad to have a s
12	"Subjects	""Admittance (Electricity)""	tkardos"	Science	PP	TK	Conductance and adr
13	"Subjects	""Adventure""	eclark"	Social science	TK	TK	Life or Lifestyles are
14	"Subjects	""Advertising and children""	pperez"	Business	AW	TK	change RT Advertisin
15	"Subjects	""Advertising effectiveness""	pperez"	Business	AW	TK	change RT Advertisin
16	"Subjects	""Age determination (Biology)""	pperez"	Science	PP	TK	Okay as is
17	"Subjects	""Age differences (Psychology)""	pperez"	Social science	TK	TK	BT Developmental ps

Orphan Term Project Methodology

Goal: Create broader term relationships to existing terms, if it complies with ANSI/NISO rules.

If not…

- Creating a new broader term is possible, but must follow policies of justification for creating new terms: usage warrant, authoritative source(s), and practicality of a new term.
- Leaving a term as an orphan is OK, but then at least an RT relationship should be present, ideally more than one.
- Changing or deleting the term (or subsuming into an existing term) might also be considered, upon further research. Occasionally, orphans are simply not good terms.

Gale Orphan Term Review Project

Orphan Term Project Methodology

Resolutions indicated on spreadsheet and entered in thesaurus management system

	A	C	D	E	F	H
1	**DESCRIPTOR**	**LAST_MOD_DA**	**LAST_MOD**	**Category**	**Owr**	**Resolution/Comments**
2	Abstracts	10/6/2005	mwendt	Humanities	HH	BT Literature
3	Access for the disabled	5/10/2010	hperry	Health/Med	HH	Okay as is
4	Albiglutide	4/23/2014	sedgar	Health/Med	HH	BT Hypoglycemic agents
5	Albinterferon alfa-2b	11/9/2010	hperry	Health/Med	HH	BT Fusion proteins
6	Allostasis	7/1/2010	dhursh	Health/Med	HH	Okay as is
7	Allostatic load	7/2/2010	dhursh	Health/Med	HH	Okay as is
8	Alterity	6/6/2011	pperez	Humanities	HH	BT Critical theory
9	Amateurism	3/17/2008	CatCleanu	Humanities	HH	Okay as is
10	Amotivational syndrome	2/17/2011	hperry	Health/Med	HH	Deactivated - concept has been debunked in medical studies and only 8 hits
11	Ancients and moderns	2/3/2009	eclark	Humanities	HH	Okay as is
12	Anglican churches	4/2/2014	rewing	Humanities	HH	BT Protestant churches
13	Antimaterialism	3/10/2008	CatCleanu	Humanities	HH	Antimaterialism (Economics) as an NPT to Anti-consumerism.
14	Apathy	4/4/2012	mobrien	Humanities	HH	BT Emotions

Causes of Orphan Terms

- During previous project that put the Subjects into a thesaurus format (changing *See also* relationships to either BT/NT or to RT as appropriate), if terms were left as orphans, that was ignored.
- Quickly created terms for immediate indexing needs, whose relationships were not completed.
- Terms for which a broader term is uncertain, and it would take time and effort, and perhaps changes to other terms (disambiguation) to resolve.
- Terms correctly created, with all correct relationships, for which a correct broader term simply does not exist.

Issues in Finding Parents to Orphan Terms

Finding imperfect Broader Terms

Stretching the permissibility of BT/NT rules

Example orphan terms and their proposed questionable broader terms:

- Atmospheric composition BT Atmosphere?
- Atmospheric haze BT Atmosphere?
- Conflict termination (Military science) BT Wars?
- Behavior problems BT Behavior?

➢ Probably OK

Issues in Finding Parents to Orphan Terms

Finding imperfect Broader Terms

Stretching the permissibility of BT/NT rules: Topics within a field
Considering the narrower term “is in” the field.

Example orphan terms and their proposed questionable broader terms:

- Coping (Psychology) BT Psychology?
- Convergence (Mathematics) BT Mathematics?
- Decision analysis BT Management science?
- Cell population BT Cytology?
- Chemical models BT Chemistry?
- Carbon rationing BT Environmental economics?

➢ Maybe not OK. (Would be OK in a hierarchical taxonomy, rather than a thesaurus.)

Issues in Finding Parents to Orphan Terms

Leaving terms as orphans (although including RTs)

Logical parent would be too broad

- College applications - We won't create BT Applications
- Animal tracks – We won't create BT Tracks

Abstract terms without broader terms

- Controversy

Complex concepts that are not what they seem

- Haunted houses – It does not belong as a narrower term to Housing (UF Houses)

Legacy LC pre-coordinated concepts that have no single broader term

- Computers and children not NT to either computers or children

Issues in Finding Parents to Orphan Terms

Parents Found!

Examples

- Alteration (Clothing) BT Tailoring
- Apathy BT Emotions
- Conscious sedation BT Anesthesia
- Stockrooms BT Storage (Physical)

Questions/Contact

Heather Hedden
Senior Vocabulary Editor
Indexing & Vocabulary Services
Metadata Standards and Services
Gale | Cengage Learning
20 Channel Center St., Boston, MA 02210
(o) 617-757-8211 | (m) 978-467-5195
heather.hedden@cengage.com
www.gale.com
www.cengage.com

heather@hedden.net
www.accidental-taxonomist.com