

# Taxonomies & Thesauri in Action

## Different Controlled Vocabularies for Different Purposes

Heather Hedden  
Senior Vocabulary Editor  
Gale, A Cengage Company

American Society for Indexing Conference  
Portland, Maine  
June 17, 2017

# About

## Heather Hedden

- Senior vocabulary editor, Gale/Cengage, 1996-2004, 2014-present
- Author of *The Accidental Taxonomist* (2010, 2016)
- Online course instructor “Taxonomies & Controlled Vocabularies”
- American Society for Indexing board member, 2016 – present
- Founder/former manager ASI Taxonomies & Controlled Vocabularies SIG
- Experienced taxonomy consultant and indexer

# Agenda

1. Introduction
  - Different purposes for Cengage controlled vocabularies
2. Comparisons of Taxonomies and Thesauri
  - Definitions, examples, and comparisons
  - Comparisons of suitability
  - When it's a compromise / in-between
3. Cengage's use of a Thesaurus and of Taxonomies
  - Views of the thesaurus and taxonomies
  - Why the difference between thesaurus and taxonomies
  - Extent of the difference between thesaurus and taxonomies / compromise
  - Meeting in the middle: links across vocabularies

# Introduction

# Introduction

## Cengage

- Publishing company of educational and reference content
- Formerly Thomson Learning, spun off from Thomson and renamed in 2007
- Two different, but related businesses:
  1. **Gale, a Cengage Company** – reference/research content:
    - Library subscription databases of periodical articles, reference book chapters/essays, digital archives, and audio/video files
    - Public web products: Questia, Books & Authors, HighBeam Research, Encyclopedia.com
    - Gale Research reference books, directories, and other book imprints
  2. **The rest of Cengage** – educational/textbook/ebook and related software publishing:
    - Former Houghton-Mifflin College Division, Wadsworth, Delmar, Heinle, Schirmer, etc.
    - Imprint brands include: 4LTR Press, Milady, Cilton, etc.
    - Companion course software (MindTap, WebAssign, Aplia, CNow, SAM, etc.)



# Introduction: Different Needs

## 1. Controlled Vocabularies for Gale Digital Products

- **Goal:** Support efficient and effective search and retrieval of items from within content repositories
  - **Design Strategy:** ANSI/NISO Z39.19 thesaurus-compliant vocabularies aimed at Boolean queries & browse navigation
  - **Content:** articles from academic journals, magazines, newspapers, reference books; some multimedia
  - **Users:** Library patrons, students, reference and research librarians,
  - **Purpose:** retrieve sufficient relevant content on a subject
- Typically, relevancy & precision takes precedence over recall.  
(Retrieving accurate results, not necessary all results)

# Introduction: Different Needs

## 2. Controlled Vocabularies for Higher Education Digital Learning Products

- **Goals:** Wide variety of specifics, but all are focused on supporting software features and functions
  - **Design Strategy:** Taxonomies with flexibility to meet specific needs of specific disciplines
  - **Content:** text, multimedia, assessments, activities, flashcards, learning objectives/outcomes, etc.
  - **Users:** Instructors, students, Cengage internal staff
  - **Purpose:** various
    - Linking of content, activities, learning objectives, etc.
    - Building custom courses (by instructors)
    - Exploration of topics in more depth (by students)
- Full recall, with more limited content, is desired.

# Comparisons of Taxonomies and Thesauri

# Comparison of Taxonomies and Thesauri

## Summary of Controlled Vocabulary Types



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

# Comparison of Taxonomies and Thesauri

Definitions – according to ANSI/NISO Z39.19 standards

## Taxonomy

A taxonomy is a controlled vocabulary consisting of preferred terms, all of which are connected in a **hierarchy** or polyhierarchy.

## Thesaurus

A thesaurus is a controlled vocabulary arranged in a known order and structured so that the various **relationships among terms** are displayed clearly and identified by **standardized** relationship indicators.

Its purpose is to promote consistency in the **indexing** of content objects...

*Sections 4.1 Definitions, 5.4.3 Taxonomy, and 5.4.4 Thesaurus*

# Comparison of Taxonomies and Thesauri

## Taxonomies

- All terms belong to a limited number of major hierarchies (or facets)
- May bend ANSI/NISO hierarchical rules.
- Supports classification, categorization, concept organization. (Like Linnaean taxonomy.)
- Approach is a top-down navigation.
- Especially serving end-users when browsing.

## Thesauri

- All terms have relationships, but “hierarchies” may be as few as 2 terms.
- ANSI/NISO rules are strictly followed.
- Supports concept scoping, disambiguation, and relationships with similar concepts. (Like Roget’s.)
- Approach is term-centered and what terms are linked to/from it.
- Especially serving indexers/ indexing.

# Comparison of Taxonomies and Thesauri: Suitability

## Taxonomies are for:

- Content and terms that naturally can be categorized
- A subject area with defined scope and limits
- Browse navigations
- Non-expert users, who benefit from guidance of hierarchies
- Relatively small collections of terms

## Thesauri are for:

- Terms that are not easily categorized into hierarchies or facets
- Multiple, overlapping subject areas or domains with diverse content.
- Highly specific terms for detailed indexing
- Users who are subject-matter experts and will likely look for specific terms
- Vocabulary that is large and/or constantly growing

# Taxonomy/Thesaurus Compromise

1. You want features of both taxonomies and thesauri.

Such as structured as hierarchies or facets *and* also associative (related term) relationships.

2. Owners or users prefer to call it one thing, even though the other is needed.

Such as users referring to a “taxonomy” for research content that needs a thesaurus: varied and broad in scope and manually indexed.

And “thesaurus management” software is for both/either.

# Cengage's use of a Thesaurus and Taxonomies

# Cengage Controlled Vocabularies

## 1. “Controlled vocabularies”

For research/library products (Gale), comprising:

- A single **Subject thesaurus**, ANSI/NISO standards compliant.  
Around 60,000 preferred terms and always growing
- Multiple (11) named entity authority files  
(People, Companies, Organizations, Geos, Works, etc.)
- Other classification metadata  
(Dates, Languages, Information Types, Genres, NAICS, etc.)

## 2. Multiple (42) “Taxonomies”

For educational/curriculum products:

- One for each academic discipline
- Ranging from 400 – 4500 preferred terms in each



## Academic Journals (201)

**Search Terms:** *Subject Guide Search (Hijab (Custom))*

Sort by: Newest

Full-text: *ARTICLE*

### [Army now allows soldiers to wear turbans, beards, and headscarves](#)

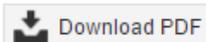
Adelle M. Banks.

The Christian Century.

134.4 (Feb. 15, 2017) p12. *Word Count: 451. Reading Level (Lexile): 1440.*

New army regulations will allow Sikh and Muslim soldiers to wear turbans, beards, and hijabs—the headscarves worn by some Muslim women—under most circumstances. "Based on the successful examples of soldiers currently...

[Link to CDL Materials Catalog](#) [ILL](#) [My Library](#) [OpenURL Link](#)



Download PDF



Save

Full-text: *ARTICLE*

### [Chancellor Angela Merkel of Germany calls for a ban of burqas as election nears](#)

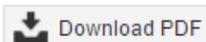
David Iaconangelo.

The Christian Century.

134.1 (Jan. 4, 2017) p15. *Word Count: 558. Reading Level (Lexile): 1320.*

In accepting her party's nomination for a fourth term in office, German chancellor Angela Merkel delivered a speech that took aim at conservative critics of her liberal asylum policies, yet also embraced a proposal to...

[Link to CDL Materials Catalog](#) [ILL](#) [My Library](#) [OpenURL Link](#)



Download PDF



Save

Abstract: *REPORT*

### [On: The couch and the chador](#)

## Content Types

**Academic Journals (201)**

Magazines (109)

Books (1)

News (2,067)

Images (10)

Videos (4)

## Search within results

Search within results

## Limit Search by:

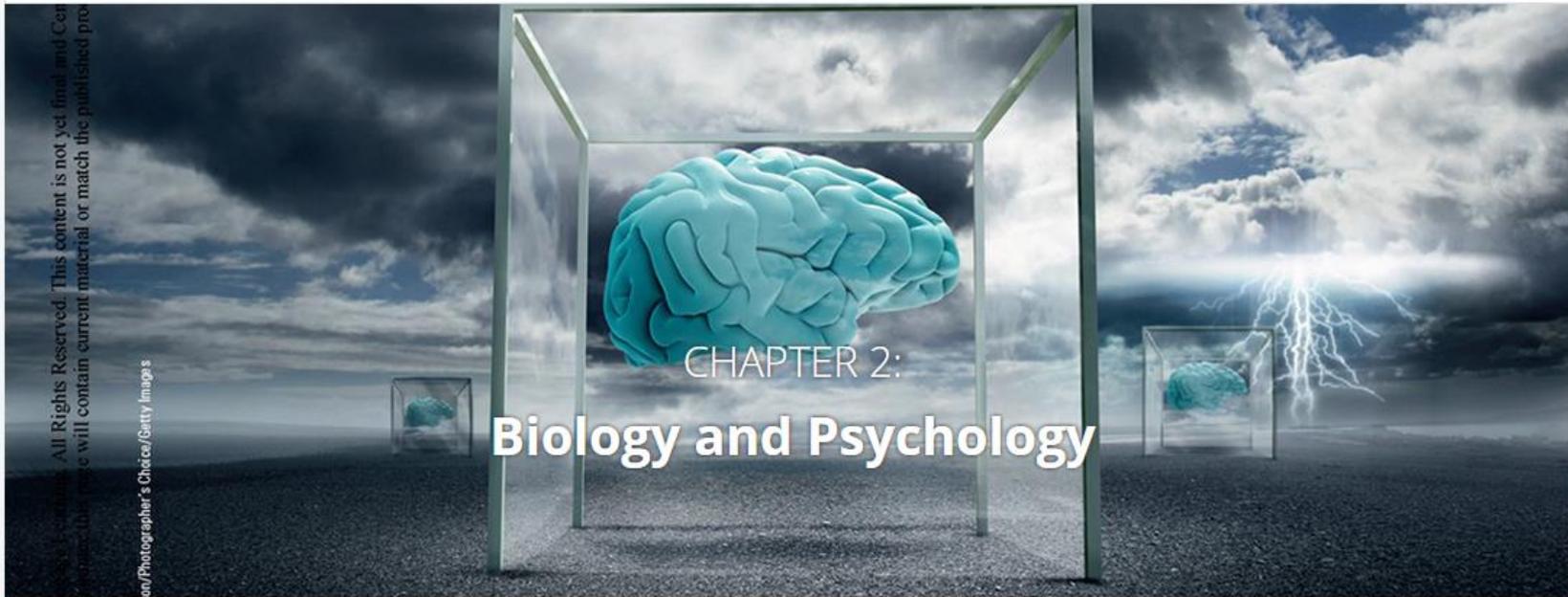
- Full Text
- Peer Reviewed Journals
- Contains Images

## Publication Dates:

- Past Week
- Past Month
- Past Year
- Custom Date Range

## Subjects

- Hijab (Custom) (201)
- Islamic clothing (54)
- Muslim women (50)



## 2-1 THE NERVOUS SYSTEM: ON BEING WIRED

The nervous system is a system of nerves involved in thought processes, heartbeat, visual-motor coordination, and so on. The nervous system consists of the brain, the spinal cord, and other parts that make it possible for us to receive information from the world outside and to act on the world. It is composed of cells, most of which are *neurons*. Here we begin our study of the nervous system.

### 2-1a. Neurons: Into the Fabulous Forest

Within our brains lies a fabulous forest of nerve cells, or neurons. [Neurons](#)

Create StudyBit

NOTES

MY UNDERSTANDING

WEAK FAIR STRONG

TAGS

Nervous system

+ Add Tags

Something About the Author, vol. 307 2017APR26 **Perms:** REFERENCE Article 1 of 1 ◀ | ▶

(Issue: -Vol: -ArticleID:490590798)

Atinuke.



Options Citation **Index** Review

**FullText** AuthAbstract FT/Abs Split English

Validate Override Add Update Delete Detail Clear

D ▼ storyteller

A S H Subfield

TSD

B Atinuke

Alpha Browsing - Google Chrome

mifhgpmsprdap1.gpms.cengage.info:7003/Per\_iMania/brAbrowse.jsp?term

**Storytellers**

- BROADER TERM:** Entertainers
- NARROWER TERM:** Griots
- SEE ALSO:** Storytelling

Storytelling

- BROADER TERM:** Oral interpretation
- SEE ALSO:** Children's stories
- SEE ALSO:** Frame narratives
- SEE ALSO:** Frame-stories
- SEE ALSO:** Story within a story
- SEE ALSO:** Storytellers

caching  periodical vocab only

**status:** ready

Atinuke

Personal

Born in Ibadan, Nigeria; daughter of a professor and an editor; married; children: two sons. Education: Attended Oxford University.

Addresses

Home--Wales.

Career

Storyteller and writer. Worked variously as a portrait painter, gardener, and artist's model.

Awards, Honors

Frankford Boase Award shortlist, 2008, and Best Children's Books selection, Bank Street College of Education, Choices selection, Cooperative Children's Book Center (CCBC), and Boston Globe/Horn Book Award Honor Book selection, all c. 2011, all for Anna Hibiscus; Best Children's Books selection, Bank Street College of Education, and CCBC Choices selection, both c. 2010, both for Hooray for Anna Hibiscus!, both 2011, for both Good Luck, Anna Hibiscus! and Have Fun, Anna Hibiscus!; CCBC Choices selection, 2011, for Anna Hibiscus's Song; 100 Titles for Reading and Sharing inclusion, New York Public Library, Best Children's

# Cengage Controlled Vocabularies

## Item Summary

**Descriptor** Hijab (Custom)

**Object** Subjects

**Categories** Anthropology, GIC Query Term, Humanities, Religion/Philosophy, Social Sciences, Tuned

**Status** Active; Approved; Preferred; Locked

**UID** 13730273

**Created** rewing 11/12/2003 07:04:24 PM

**Modified** ddigregio 11/12/2015 06:43:23 PM

**Sort Key** hijab custom

Save

Delete

Combine

Refresh

Deactivate

Make NPT

## Descriptor Elements

Descriptor: Hijab (Custom)

## Extended Attributes

Scope:

Source: PerLit; WWW usage (Google); LCSH\$ (Purdah); EB\$ (Purdah)

Islamic and Hindu custom of women being fully covered with clothing when outside

## [Add Relationship]

## Relationships

[UF Purdah \(Custom\) \(Subjects\)](#)

[BT Manners and customs \(Subjects\)](#)

[RT Dress codes \(Subjects\)](#)

[RT Hindu women \(Subjects\)](#)

[RT Islamic clothing \(Subjects\)](#)

[RT Islamic culture \(Subjects\)](#)

[RT Muslim women \(Subjects\)](#)

[RT Veils \(Subjects\)](#)

[EQTAXON Islamic clothing \(Taxon Religion\)](#)

[EQD Hijab \(Custom\) \(FCDTerms\)](#)

[IDX ai156vi0012 The French Headscarf Ban Is Oppressive \(Doc\)](#)

[IDX ai156vi0013 Fabric of Society: Banning Headscarves Is Right \(Doc\)](#)

[IDX GIC2008ov\\_000068 Headscarves: Religious and Political Issues \(Doc\)](#)

[IDX GIC2010stat\\_001701 G1278xb02 International Views of Bans on Full Islamic Veils: 2010 \(IM\)](#)

[IDX GICTPC1278 Headscarves: Religious And Political Issues \(Portals\)](#)

[IDX HID0416 Hijab and Islam \(Doc\)](#)

# Cengage Controlled Vocabularies: Why the Difference

1. Why different controlled vocabularies for research products and learning/educational products?
2. Why taxonomies vs. a thesaurus?

# Cengage Controlled Vocabularies: Why the Difference

## Why one subject thesaurus for research products:

- Multiple, overlapping subject areas.
- Some terms cannot be easily categorized.
- Both Hierarchical (NT, BT) and Related Term (RT) relationships important for communicating meaning and disambiguation.
- Terms can be broad because "sub-headings" are applied to focus meaning.
- Developed for use by trained indexers, with plenty of RTs and cross-references.
- Vocabulary is very large and growing.

# Cengage Controlled Vocabularies: Why the Difference

## Why new, distinct taxonomies for academic disciplines:

- Can be more customized around a discipline.
- May contain more, specific terms.
- Instructional vs. research perspective often calls for different terms.
- Eliminates the need to disambiguate same terms with different meanings.
- Single discipline can allow for smaller, more efficient vocabulary.
- More flexibility for uses in new products not yet identified.
- Have complex terms and no need for a set of second-level “sub-headings” vocabulary.

# Cengage Controlled Vocabularies: Why the Difference

Additional benefits of hierarchical structure taxonomies for learning disciplines:

- Facilitates limiting the taxonomy to a subject discipline scope.
- Easier to understand and review by non-taxonomist subject-matter experts.
- Categories with narrower terms are more typical in instructional material.

Such as learning objectives for “describe three types of X”,  
with X being the broader term/category for three narrower terms.

# Cengage Controlled Vocabularies: Extent of the Difference

## Taxonomies for educational content

- Separate taxonomies for each discipline (subject area).
- Term choice/form reflects the discipline. No need for parenthetical qualifiers
- Each relatively small (typically 500 – 3,000 preferred terms).
- A single hierarchy for each discipline.
- Once built and tested, vocabularies will be slow-growing.
- Easier use for non-indexers: subject-matter experts & product developers.
- Could be implemented in an end-user browse interface.

# Cengage Controlled Vocabularies: Extent of the Difference

Cengage learning discipline “taxonomies” are a hybrid:

## ■ **Thesaurus-like:**

- Have all standard thesaurus relationships
  - Broader term/narrower term
  - Related terms (although not as extensive)
  - Nonpreferred terms
  - Scope notes

## ■ **Taxonomy-like:**

- Each taxonomy has a single top term and limited number of 2<sup>nd</sup> level terms.
- Sometimes bends ANSI/NISO hierarchical relationship rules at the top levels.

# Cengage Controlled Vocabularies: Extent of the Difference

## Health Insurance

term in the Subject thesaurus

- [NT Capitated payment systems \(Medical care\) \(Subjects\)](#)
- [NT Group health insurance \(Subjects\)](#)
- [NT Hospitalization insurance \(Subjects\)](#)
- [NT Long term care insurance \(Subjects\)](#)
- [NT Managed care plans \(Medical care\) \(Subjects\)](#)
- [NT Medicare \(Subjects\)](#)
  - [NT Medicare Part D \(Subjects\)](#)
- [NT Medigap \(Subjects\)](#)
- [NT National health insurance \(Subjects\)](#)
- [NT Point of service plans \(Subjects\)](#)
- [NT Preferred provider organizations \(Medical care\) \(Subjects\)](#)
- [NT Prescription drug plans \(Subjects\)](#)
  - [NT Medicare Part D \(Subjects\)](#)
- [NT Prospective payment systems \(Medical care\) \(Subjects\)](#)
  - [NT Diagnosis related groups \(Subjects\)](#)
- [NT Relative value scale payment systems \(Medical care\) \(Subjects\)](#)
- [NT Single payer system \(Health care\) \(Subjects\)](#)
- [RT Accident insurance \(Subjects\)](#)
- [RT Altmark Trans GmbH v. Nahverkehrs-gesellschaft Altmark GmbH 2003 C.M.L.R. 12 \(E.C.J.\) \(Court Cases\)](#)
- [RT American Ass'n of Retired Persons v. EEOC 489 F.3d 558 \(3d Cir. 2007\) \(Court Cases\)](#)
- [RT Coinsurance \(Subjects\)](#)
- [RT Commission of the European Communities v. Ireland 2012 C.E.C. 229 \(C.J.E.U.\) \(Court Cases\)](#)
- [RT Cooperative insurance \(Subjects\)](#)
- [RT Cost shifting \(Medical care\) \(Subjects\)](#)
- [RT Dental insurance \(Subjects\)](#)
- [RT Diagnosis related groups \(Subjects\)](#)
- [RT Disability insurance \(Subjects\)](#)
- [RT Flexible spending accounts \(Subjects\)](#)
- [RT Fraternal organizations \(Subjects\)](#)
- [RT Germany. Sickness Insurance Act 1883 \(Statutes\)](#)
- [RT Health care costs \(Subjects\)](#)
- [RT Health care industry \(Subjects\)](#)
- [RT Health care reform \(Subjects\)](#)
- [RT Health insurance claims processing software \(Subjects\)](#)
- [RT Health insurance exchanges \(Subjects\)](#)
- [RT Health insurance industry \(Subjects\)](#)
- [RT Health insurance tax credits \(Subjects\)](#)
- [RT Health maintenance organizations \(Subjects\)](#)

4 UFs  
14 direct NTs  
(16 NTs total)  
up to 2 levels deep  
36 RTs

## Health Insurance

term in the Health discipline taxonomy

- [NT Health care reimbursement systems \(Taxon Health\)](#)
  - [NT CMS reimbursement \(Taxon Health\)](#)
  - [NT Health insurance coverage \(Taxon Health\)](#)
  - [NT Health insurance history \(Taxon Health\)](#)
- [NT Health insurance laws \(Taxon Health\)](#)
  - [NT ENTITY Health Care and Education Affordability Reconciliation Act of 2010 \(Statutes\)](#)
  - [NT ENTITY Patient Protection and Affordable Care Act \(Statutes\)](#)
  - [NT Health insurance terminology \(Taxon Health\)](#)
- [NT Insurance claims process \(Taxon Health\)](#)
  - [NT Accept assignment \(Taxon Health\)](#)
  - [NT Assignment of benefits \(Taxon Health\)](#)
  - [NT Claims adjudication \(Taxon Health\)](#)
  - [NT Claims tracking \(Taxon Health\)](#)
  - [NT Electronic data interchange \(EDI\) \(Taxon Health\)](#)
- [NT Filing insurance claims \(Taxon Health\)](#)
  - [NT CMS-1500 claim form completion \(Taxon Health\)](#)
  - [NT National provider identifier \(NPI\) \(Taxon Health\)](#)
  - [NT Payer responsibility determination \(Taxon Health\)](#)
  - [NT Insurance claim files maintaining \(Taxon Health\)](#)
- [NT Insurance claim forms \(Taxon Health\)](#)
  - [NT CMS-1500 claim forms \(Taxon Health\)](#)
    - ... More levels exist.
    - [NT UB-04 claim forms \(Taxon Health\)](#)
- [NT Insurance forms \(Taxon Health\)](#)
  - [NT First report of injury form \(Taxon Health\)](#)
- [NT Insurance claim forms \(Taxon Health\)](#)
  - [NT CMS-1500 claim forms \(Taxon Health\)](#)
    - ... More levels exist.
    - [NT UB-04 claim forms \(Taxon Health\)](#)
- [NT Insurance payments \(Taxon Health\)](#)
- [NT Insurance plan types \(Taxon Health\)](#)
  - [NT Commercial insurance plans \(Taxon Health\)](#)
    - [NT Automobile, disability, and liability insurance \(Taxon Health\)](#)
    - [NT Medigap \(Taxon Health\)](#)
    - [NT Consumer-driven health plans \(Taxon Health\)](#)
    - [NT Fee-for-service insurance \(Taxon Health\)](#)
  - [NT Government health plans \(Taxon Health\)](#)
  - [NT Medicaid \(Taxon Health\)](#)
    - ... More levels exist.

3 UFs  
13 direct NTs  
(58 NTs total)  
up to 4 levels deep  
5 RTs

## Cengage Controlled Vocabularies: Meeting in the Middle

Discipline taxonomy terms are linked to equivalent Subject Thesaurus terms.

- Vocabulary management system “Atlas” supports customized cross-vocabulary equivalency links (crosswalks).
- When implemented in products, students can jump from an instructional topic to relevant articles in the Gale reference databases, indexed with the equivalent term.

Item Summary		[Add Relationship]
<b>Descriptor</b>	Pressure caps	<b>Relationships</b> <a href="#">UF Radiator caps (Taxon Auto)</a> <a href="#">BT Cooling systems and equipment (Taxon Auto)</a> <a href="#">RT Radiators (Taxon Auto)</a> <a href="#">EQTAXON Pressure caps (Vehicles) (Subjects)</a>
<b>Object</b>	Taxon Auto	
<b>Categories</b>		
<b>Status</b>	Active; Approved; Preferred; Unlocked	
<b>UID</b>	17983663	
<b>Created</b>	mobrien 11/02/2006 01:00:17 PM	
<b>Modified</b>	tkardos 01/06/2015 12:50:47 PM	
<b>Sort Key</b>	pressure caps	

# Cengage Controlled Vocabularies: Meeting in the Middle

Discipline taxonomy terms are linked to selected narrower named entities in the Gale name authority files.

- “Atlas” supports customized cross-vocabulary hierarchical relationships.
- Eliminates the need for duplicate maintenance.

The screenshot displays the 'Item Summary' and 'Relationships' sections of the Cengage Atlas interface. The 'Item Summary' section on the left contains the following information:

- Descriptor:** Pharmacy laws and regulations
- Object:** Taxon Health
- Categories:** (empty)
- Status:** Active; Approved; Preferred; Unlocked
- UID:** 22817447
- Created:** hhedden 11/11/2015 03:59:02 PM
- Modified:** hhedden 05/31/2016 08:48:43 PM
- Sort Key:** pharmacy laws and regulations

Below the summary are buttons for 'Save', 'Delete', 'Combine', 'Refresh', 'Deactivate', and 'Make NPT'. The 'Descriptor Elements' section at the bottom shows the descriptor 'Pharmacy laws and regulations' in a text box.

The 'Relationships' section on the right, titled '[Add Relationship]', lists the following relationships:

- [UF Pharmaceutical laws \(Taxon Health\)](#)
- [UF Pharmacy regulation \(Taxon Health\)](#)
- [BT Law in health care \(Taxon Health\)](#)
- [BT Pharmacy practice \(Taxon Health\)](#)
- [NT ENTITY Dietary Supplement Health and Education Act of 1994 \(Statutes\)](#)
- [NT ENTITY Drug Listing Act of 1972 \(Statutes\)](#)
- [NT ENTITY Poison Prevention Packaging Act of 1970 \(Statutes\)](#)
- [NT ENTITY Ryan Haight Online Pharmacy Consumer Protection Act of 2008 \(Statutes\)](#)
- [RT Controlled substances \(Taxon Health\)](#)
- [RT Drug laws and regulations \(Taxon Health\)](#)
- [RT Federal regulations \(Taxon Health\)](#)

The 'NT ENTITY Dietary Supplement Health and Education Act of 1994 (Statutes)' relationship is circled in red.

# Summary

1. For multiple purposes (different content and different users) a single controlled vocabulary may not suffice, even within the same organization.
2. For different purposes sometimes a thesaurus is needed, and sometimes a taxonomy is needed.
3. The distinction between a thesaurus and a taxonomy can be blurred.
4. Multiple controlled vocabularies can be linked term-by-term.

# Questions/Contact

## Heather Hedden

Senior Vocabulary Editor  
Gale, A Cengage Company  
20 Channel Center Street  
Boston, MA 02210

[heather.hedden@cengage.com](mailto:heather.hedden@cengage.com)  
[heather@hedden.net](mailto:heather@hedden.net)

[www.gale.com](http://www.gale.com)

[www.cengage.com](http://www.cengage.com)

<http://accidental-taxonomist.blogspot.com>