

Data, Data Everywhere: How Information Professionals Categorize and Systemize the Data

Taxonomy Projects for Internal Content Management

by

Heather Hedden

Taxonomy Consultant, Hedden Information Management

SLA@Simmons Student Group, SLA New England Chapter

Simmons University, Boston

February 12, 2020

About Heather Hedden

- Taxonomy consultant
 - Independent, through Hedden Information Management
 - Previously as an employed and contract consultant
- Former staff taxonomist
 - At various companies: Gale/Cengage Learning, Viziant, First Wind
- Instructor of online and onsite taxonomy courses
 - Independently through Hedden Information Management
 - Previously at Simmons University - Library & Information Science School CE
- Author of *The Accidental Taxonomist* (2010, 2016, Information Today, Inc.)
- Member of SLA and its Taxonomy Division since 2009

- Started career as periodical indexer for a library database vendor

What is a Taxonomy?

- A kind of controlled vocabulary or knowledge organization system (has designated preferred terms for each unambiguous concept)
- Terms are used to tag/index/categorize content items to make them easier to retrieve, supporting better precision and recall than search alone
- Terms are arranged in a structure of hierarchies, categories, or facets
- Optional: Terms may have alternative labels (synonyms/nonpreferred terms) to redirect to user to the preferred term.
- Optional: Terms may have additional kinds of relationships, notes, and attributes

What is a Taxonomy?

Taxonomy Examples

Leisure and culture

- . Arts and entertainment venues
 - . Museums and galleries
- . Children's activities
- . Culture and creativity
 - . Architecture
 - . Crafts
 - . Heritage
 - . Literature
 - . Music
 - . Performing arts
 - . Visual arts
- . Entertainment and events
- . Gambling and lotteries
- . Hobbies and interests
- . Parks and gardens
- . Sports and recreation
 - . Team sports
 - . Cricket
 - . Football
 - . Rugby
 - . Water sports
 - . Winter sports
- . Sports and recreation facilities
- . Tourism
 - . Passports and visas
- . Young people's activities

Hierarchical Taxonomy Example

Career Level

- Student
- Entry Level
- Experienced
- Manager
- Director
- Executive

Function

- Customer Service & Support
- Delivery
- Engineering
- Finance
- General Management
- Legal & Regulatory Affairs
- Marketing & Advertising [more]

Industry

- Agriculture
- Apparel & Fashion
- Automotive
- Aviation & Aerospace
- Banking
- Biotechnology
- Broadcast Media
- Chemicals [more]

Faceted Taxonomy Example

What is a Taxonomy?

Taxonomy vs. a Classification System

Classification systems: LC Classification, Dewey Decimal Classification, NAICS, etc.

Classification Systems

- Have codes
- Don't have synonyms (alternate labels)
- Comprehensive and balanced coverage of a domain
- Designed to be browsed hierarchically, top-down
- Built for a domain but independent of specific content
- Provide for limited expansion, within the structure

Taxonomies

- Don't have codes
- Often have synonyms (alternate labels)
- Just the terms needed and may be hierarchically unbalanced
- Designed to be browsed, searched, or may not be fully displayed to end-users
- Custom-built to include the topics in the content to be tagged
- Can grow and adapt without limits

What is a Taxonomy?

Common Knowledge Organization System Types



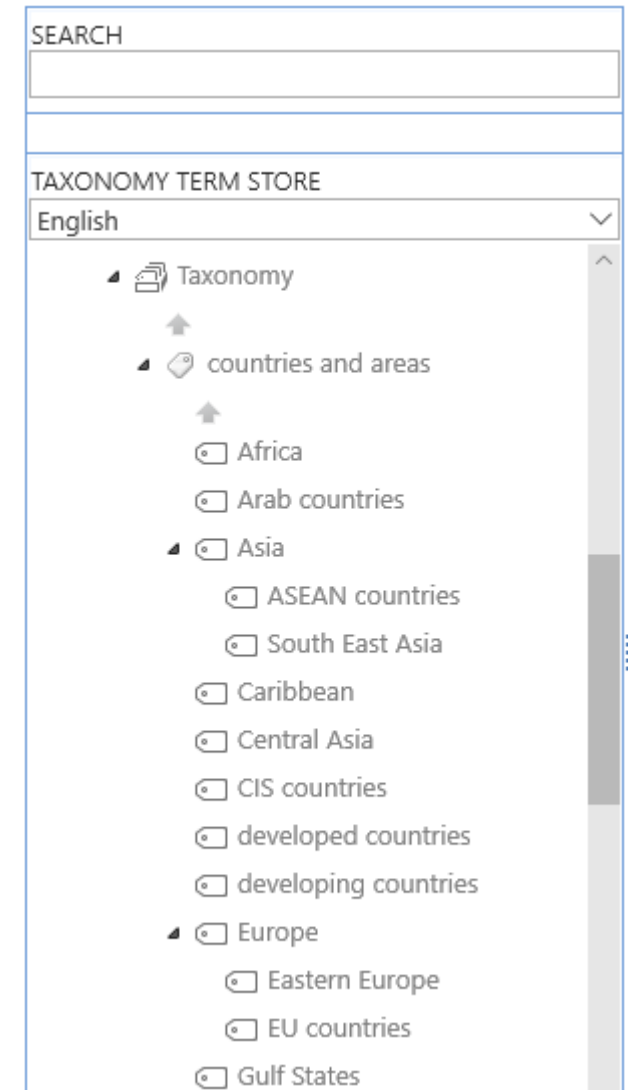
Term List	Synonym Ring	Authority File	Taxonomy	Thesaurus	Ontology
Ambiguity control	Synonym control	Ambiguity control Synonym control	Ambiguity control (Synonym control) Hierarchical relationships	Ambiguity control Synonym control Hierarchical relationship Associative relationships	Ambiguity control (Synonym control) Semantic relationships Classes Attributes (Linked data)

Taxonomy Creation for Internal Content Management: Challenges

- A single shared taxonomy across the entire enterprise supporting searching on everything, *or* multiple site/department-specific taxonomies more tailored to specific content repositories.
 - Advantage/disadvantages to both approaches – which to do?
- Content owners, as stakeholders, contribute term ideas to the taxonomy
 - They tend to think of classification and not of search terms to limit results
- Difficulties distinguishing document types from topics in some cases
 - E.g. *Doc Type: Evaluation report* or *Doc Type: Report + Topic: Evaluation*
- Content management system limitations on taxonomy features
 - E.g. no polyhierarchy (a term cannot have more than one broader term)
- People who tag documents are content owners, not information professionals
 - Require training
 - May not like to tag

Taxonomy Projects for SharePoint

- Companies/organizations adopt Microsoft SharePoint for varied reasons. (Intranet, enterprise content management, collaboration space, etc.)
- Documents to share can be posted and tagged to be searched and found.
- Tags can be either shared uncontrolled keywords or controlled taxonomies.
- Taxonomy is managed in SharePoint's "Term Store" with:
 - Uniqueness enforced
 - Hierarchies (hierarchical relationships)
 - Synonyms ("other labels"/ nonpreferred terms)
 - Scope notes / description
 - Searchable and browsable



Taxonomy Projects for SharePoint: Case Studies

International Labour Organization

- Migrating content from an intranet platform that didn't have search (so had no taxonomy) to SharePoint as its intranet.
- Has a taxonomy for its public website that they preferred to expand to cover internal content, rather than keep as separate taxonomies
 - Much of the public website taxonomy is not appropriate for the SharePoint content and its terms could be confusing to those tagging **Health** and **Health policy** (external topics) vs. **Health and wellbeing** (internal human resources topic)
- Taxonomy term input from content owners tended to focus on classification of functions not on topics of documents (e.g. proposed term for **Editing**)

Taxonomy Projects for SharePoint: Case Studies

Small professional financial services firm

- Maintains all corporate and financial documents for its clients, migrating documents from shared nested folders in box (cloud content management) to SharePoint
- Has deep hierarchy of folder categorization, with some inconsistencies across clients, and empty folders
 - e.g. **P&L** and **Balance sheets** in both **Monthly close** folder for one client and **Reports and Analysis > GL > Management Reports** folder for another client.
- Folder names mix document types, topics, and purposes
 - Propose 3 facets:
 - Doc Type:* **Balance sheet**
 - Topic:* **P&L**
 - Purpose:* **Monthly close; Management reporting**

Further Reading

- Classification vs. taxonomy of search terms
 - Blog post: “Classification Systems vs. Taxonomies” (February 9, 2020)
<https://accidental-taxonomist.blogspot.com/2020/02/classification-systems-vs-taxonomies.html>
- A single shared taxonomy across the entire enterprise supporting searching on everything, *or* multiple site/department-specific taxonomies
 - Blog post: “Intranet and ECM Taxonomies” (January 13, 2020)
<http://accidental-taxonomist.blogspot.com/2020/01/intranet-and-ecm-taxonomies.html>
- SharePoint not permitting polyhierarchy in taxonomies
 - Blog post: “Polyhierarchy in the SharePoint Term Store” (January 30, 2016)
<http://accidental-taxonomist.blogspot.com/2016/01/polyhierarchy-in-sharepoint-term-store.html>

Questions/Contact

Heather Hedden
Hedden Information Management

heather@hedden.net

98 East Riding Dr.
Carlisle, MA 01741
978-467-5195

www.hedden-information.com

accidental-taxonomist.blogspot.com

twitter.com/hhedden

[linkedin.com/in/hedden](https://www.linkedin.com/in/hedden)