



# Building Taxonomies from the Bottom Up and Top Down Technologies and Collaboration

Heather Hedden July 16, 2024 Kingston, RI



Heather Hedden Senior Consultant Enterprise Knowledge, LLC (1)

- Providing taxonomy and related consulting services to diverse corporate, government, and nonprofit clients
- In taxonomist roles for 28 years, as a consultant and staff taxonomist
- Previously a controlled vocabulary editor with Gale/Cengage
- Author of the book *The Accidental Taxonomist*, 3<sup>rd</sup> edition
- Instructor of taxonomy workshops and courses



## Enterprise Knowledge at a Glance **RC**

**ESTABLISHED 2013** – OUR FOUNDERS AND PRINCIPALS HAVE BEEN PROVIDING KNOWLEDGE MANAGEMENT CONSULTING TO GLOBAL CLIENTS FOR OVER 20 YEARS.

# **AREAS OF EXPERTISE**

KM STRATEGY & DESIGN
 TECHNOLOGY SOLUTIONS
 CONTENT & BRAND STRATEGY
 ENTERPRISE SEARCH
 ENTERPRISE LEARNING

## STABLE CLIENT BASE



#### C TAXONOMY & ONTOLOGY DESIGN

- CAGILE, DESIGN THINKING, & FACILITATION
- C KNOWLEDGE GRAPHS, DATA MODELING, & AI
- CINTEGRATED CHANGE MANAGEMENT
- CONTENT MANAGEMENT

#### HEADQUARTERED IN WASHINGTON, DC,



# CONSULTANTS

#### KMWORLD'S

100 COMPANIES THAT MATTER IN KM (2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024)

TOP 50 TRAILBLAZERS IN AI (2020, 2021, 2022)

#### **CIO REVIEW'S**

20 MOST PROMISING KM SOLUTION PROVIDERS (2016)

#### INC MAGAZINE

#2,343 OF THE 5000 FASTEST GROWING COMPANIES (2021) #2,574 OF THE 5000 FASTEST GROWING COMPANIES (2020) #2,411 OF THE 5000 FASTEST GROWING COMPANIES (2019) #1,289 OF THE 5000 FASTEST GROWING COMPANIES (2018)

#### INC MAGAZINE

BEST WORKPLACES (2018, 2019, 2021, 2022)

#### WASHINGTONIAN MAGAZINE'S

TOP 50 GREAT PLACES TO WORK (2017)

#### WASHINGTON BUSINESS JOURNAL'S

BEST PLACES TO WORK (2017, 2018, 2019, 2020)

#### ARLINGTON ECONOMIC DEVELOPMENT'S

FAST FOUR AWARD FASTEST GROWING COMPANY (2016)

#### VIRGINIA CHAMBER OF COMMERCE'S

FANTASTIC 50 AWARD – FASTEST GROWING COMPANY (2019, 2020)

# Outline

- Introduction to taxonomies
- Taxonomy design & creation
- Top-down and user-focused tasks
- Bottom-up and content-focused task
- Conclusions



knowledge	Search	Agile, Design Thinking, Facilitation (60)
Knowledge management	eBay ) Electropics ) Comerce & Ph	Artificial Intelligence (25
Knowledge	ebay / Electronics / Cameras of h	Change Management & Communications (23)
Knowledge-based systems	Cameras & P	Company (19)
Knowledge transfer		Content & Brand Strates
Knowledge workers	Shop by Category	Enterprise Learning (23)
Knowledge workers	Binoculars & Telescopes 🔨	Enterprise Search (49)
KnowledgeWare Inc	See all in Binoculars & Telescopes	Modeling (108)
Knowledge acquisition (Expert systems)	Binocular Cases & Accessories Binoculars & Monoculars	Knowledge Manageme Strategy & Design (243)
Knowledge Adventure Inc.	Telescope Parts & Accessories	Design (113)
Knowledge Networks Inc.	Telescopes	Technology Solutions (9
Gale Academic Onefile	Camcorders	
	Camera Drones	Article Type
еВау	Camera Manuals & Guides	Blog (324)
xonomies help people	find information	Podcast (71)
		Presentation (56)

#### ENTERPRISE KNOWLEDGE

#### Topic

Advanced Content (73)

gy (5)

ata

nt

8)

- Case Study (40)
- White Paper (31)

- Computer and information sciences

YO

#### Artificial intelligence

- · Artificial neural networks
- · Expert systems
- Genetic programming
- Machine learning

#### **Machine learning**

- Decision tree learning
- Deep learning
- Ensemble methods
- · Machine learning algorithms
- Relevance vector machines
- · Supervised machine learning
- · Support vector machines
- Unsupervised machine learning

PLOS One

Enterprise Knowledge website knowledge base



## **Taxonomy Defined**

A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy.

– ANSI/NISO Z39.19-2005 r2010 section 4.1 Definitions

## What is a taxonomy not?

- Not just any metadata or tags
- Not a business glossary
- Not a classification scheme
- Not a navigation scheme

- □ Needs to be controlled, structured, related
- □ For finding, not defining
- □ For tagging, not classifying
- For searching, too, not just browsing \$







## **Taxonomy Features**

#### It's Controlled:

A taxonomy is kind of controlled vocabulary, based on unambiguous concepts, not just words (things, not strings). A concept may have multiple labels: preferred and alternative.

#### It's Organized:

The concepts are organized in a structure of hierarchies, categories, or facets to make them easier to find and understand.



0

6

 $\sim$ 

~

 $\sim$ 

 $\sim$ 

 $\sim$ 

V



## Taxonomy **Structural** Types

Leisure and culture				Career Level
<ul> <li>Arts and entertainme</li> <li>Museums and gal</li> <li>Children's activities</li> <li>Culture and creativity</li> </ul>	nt venues leries			<ul> <li>Student</li> <li>Entry Level</li> <li>Experienced</li> <li>Manager</li> <li>Director</li> </ul>
Crafts	Hierarchi	cal		Executive
Literature	Taxonom	וע   F	Function	
Music Performing arts	Example			<ul> <li>Customer Se</li> <li>Delivery</li> <li>Engineering</li> </ul>
Visual arts				<ul><li> Finance</li></ul>
. Entertainment and ev	vents			General Man
Hobbies and interest	5 S			Legal & Regu
. Parks and gardens	<b>.</b>			
. Sports and recreation	n			[[[[0]0]]
Team sports				Industry
Cricket				Agriculture
Fooldall Rugby				<ul> <li>Apparel &amp; Fa</li> <li>Automotive</li> </ul>
Water sports				Aviation & Ae
Winter sports				Banking
. Sports and recreation	n facilities			Biotechnolog
. Tourism				Broadcast Me
. Passports and vis	as			
. roung people's activi	ties			[IIIOIC]

Faceted Taxonomy Example

- ner Service & Support
- Y
- ering
- e
- al Management
- & Regulatory Affairs
- ting & Advertising
- lture
- el & Fashion
- otive
- on & Aerospace
- g
- hnology
- cast Media
- cals

<b>000 Computer science, knowle</b> 010 Bibliographies	dge & systems		350 Public administration & military science 360 Social problems & social services	700 Arts
020 Library & information sciences			370 Education	720 Architecture
030 Encyclopedias & books of facts			380 Commerce, communications & transportation	730 Sculpture, ceramics & metalwork
040 [Unassigned]			390 Customs, etiquette & folklore 400 Language	740 Drawing & decorative arts
050 Magazines, journals & serials			400 Language	750 Painting
060 Associations, organizations & mus	eums		410 Linguistics	760 Graphic arts
070 News media, journalism & publishi	ing		420 English & Old English languages	770 Photography & computer art
080 Quotations	•	-	430 German & related languages	780 Music
090 Manuscripts & rare books	Dewey Decim	nal	440 French & related languages	790 Sports, games & entertainment
100 Philosophy	Classification		450 Italian, Romanian & related languages	800 Literature, rhetoric & criticism
110 Metaphysics	Clussification		460 Spanish & Portuguese languages	810 American literature in English
120 Epistemology	100s level		470 Latin & Italic languages	820 English & Old English literatures
130 Parapsychology & occultism			480 Classical & modern Greek languages	830 German & related literatures
140 Philosophical schools of thought			490 Other languages	840 French & related literatures
150 Psychology	ls a classificati	ion	500 Science	850 Italian, Romanian & related literatures
160 Logic	gic cyctom a		510 Mathematics	860 Spanish & Portuguese literatures
170 Ethics	systema		520 Astronomy	870 Latin & Italic literatures
180 Ancient, medieval & eastern philos taxonomy?			530 Physics	880 Classical & modern Greek literatures
190 Modern western philosophy	5		540 Chemistry	890 Other literatures
200 Religion			550 Earth sciences & geology	900 History
210 Philosophy & theory of religion			560 Fossiis & prenistoric life	910 Geography & travel
220 The Bible			570 Life sciences; biology	920 Biography & genealogy
230 Christianity & Christian theology			580 Plants (Botany)	930 History of ancient world (to ca. 499)
240 Christian practice & observance			590 Animais (20010gy)	940 History of Europe
250 Christian pastoral practice & religio	ous orders		600 Technology	950 History of Asia
260 Christian organization, social work	x & worship		610 Medicine & health	960 History of Africa
270 History of Christianity			620 Engineering	970 History of North America
280 Christian denominations			630 Agriculture	980 History of South America
290 Other religions			650 Management & public relations	990 History of other areas
300 Social sciences, sociology & anthropology			660 Chomical anginaaring	
310 Statistics			670 Manufacturing	
320 Political science	Г	_		
330 ECONOMICS		Also, d	does not reflect the context of a	a particular content set.
JHU LAW				

.

## What is a Taxonomy For?

- Concepts/terms are used to tag/index/categorize content items to make them easier to be found and retrieved
  - Supporting better findability than search alone
- The taxonomy is an intermediary that links the user to the desired content
- The taxonomy should be designed for the particular *content* and *users*
- Both content and users provide sources and input for the taxonomy





# Taxonomy Design & Creation

## Generic (Traditional) Taxonomy Design Approaches

## Bottom Up

- 1. Identify the specific concepts needed for tagging the content
- 2. Then group the concepts into categories and then broader categories, which form the hierarchies and/or facets

#### • Top Down

RPRISE KNOWLEDGE

- 1. Identify the top concepts/terms, hierarchies and/or facets needed to cover the subject domain
- 2. Then add narrower concepts, and further narrower concepts to build out the hierarchies to the level of detail desired

In combination, they can meet (and overlap) in the middle.







## Additional, Business Taxonomy Design Approaches

#### **Bottom-up Methods Focus On:**

- The concepts, terms, entities within the content that will be searched for and retrieved with the taxonomy
- Content analysis, tagging/indexing
- The back end

#### **Top-down Methods Focus On:**

- Users, use cases, the application, the subject domain scope
- Information architecture, user experience, taxonomy displays
- The front end

Both approaches are needed in combination.

In business taxonomies, top-down is particular important.

1				П	I
			22	11	I
	_			11	I
1.25	_			11	I
				11	I
	_			۱ŀ	





## Value of Business Taxonomy Design Approaches

#### Value of Bottom-up Methods:

- Alignment Taxonomy concepts are reflected and accurately distributed across content.
- Completeness The taxonomy concepts are applicable to the complete set of content across the system.

#### Value of Top-down Methods:

RPRISE KNOWLEDGE

 Usability - The structure and language of the taxonomy are intuitive to end users.



## Value of Business Taxonomy Design Approaches



## Cost savings

By ensuring the taxonomy is user-centric and applies to the complete set of content, an organization reduces the need for costly taxonomy re-work in the long-term.

## Increased user adoption

By ensuring the taxonomy reflects the language end users use, an organization lays the groundwork for **intuitive and effective search and tagging**,

ultimately increasing the likelihood and ease of adoption of the new taxonomy.

# Increased user satisfaction

By aligning the taxonomy with user needs this **improves the user experience**. End users will be **more productive and efficient**, completing daily tasks with fewer roadblocks.



# Top-Down Taxonomy Design

## **Top-Down Taxonomy Design**

Used to capture patterns of business process and user-centric understanding to make the taxonomy intuitive and user-friendly.

- Especially focused on the **users**
- Makes more use of collaboration

## Methods

- One-on-one stakeholder interviews
- Focus groups
- Asynchronous interviews, questionnaires, or surveys
- System demos
- Taxonomy brainstorming and design workshops
- Card-sorting activities

Use when creating a new taxonomy or expanding an existing one.



## **Top-Down Methods**

#### Interviews

Discussing content and taxonomy needs with specific individuals, such as key project stakeholders in a senior leadership role or people who have unique roles.



#### **Focus Groups**

Conducting taxonomy focus groups per business area to identify metadata fields that are applicable to the organization as a whole and that are unique to their own business area.



#### System Demos

Attending demos of content systems to learn more about the organization's content, search, tagging, and taxonomy needs.



JULY 14-16, 2024 // UNIVERSITY OF RHODE ISLAN

## **Top-Down Methods, Continued**

#### Workshops

Working with a cross-organizational group of stakeholders and guiding them to provide metadata and taxonomy details by asking key questions about content they create or use.



Possible workshop activities:

- Brainstorming concepts and metadata types
- Content tagging exercise
- Card sorting exercise

With online tools, activities could be asynchronous, but they are more collaborative and productive when done in a workshop together.

Put yaur id	ees in sticky n	otes below - o	ine idea per s	icky notel	Prepare to	share w	th the rest of	the workshop!		
Енатріє: Торіс	timeframe		File Format		Thi Thi (Fo 71 Mete	se of kning cma'- 105; smag	cantent type (document, image, video, etc.)	Videos		
Author	Publish Data	Other Laboratory	(nstitution (lab, rec. doe, etc.)				weste form	host rock	program (YM, WIPP, SPWST, etc.)	sensitivi
Physical Hardivare	Type of Instrument		File Size		ttle		Transportation Perkept	type of schepand Physical safepand in protect nucleor system;	system (NDS, EBS, biosphere)	OUDIOUR
		instrument Celibration	Parating Searce	keyword		tapic			iece of	
Type of Document	Analysis files Stpat, C cutpat/	MFRC cur focuments	GA nentation Facility Relea Waste	Press	2	ipe of Data	americal	FEP Screen Argune	Parometer ing kustications rets	
25.83 Te 0	ohnical' I cuments Doc	Nork ocum Untents NWDJ SJ	NY Records		Xest	dete : Se	perimental			



## **Card Sorting**

- A tool to learn how end users think content should be organized and categorized and how categories should be grouped.
- Card sorting provides insight into how people conceptualize, group, and label ideas.
- Card sorting may be used to not only design taxonomies, but also to test those taxonomies.

#### OptimalSort Shopper Bags Shoes Socks Tools & Accessories Ankle Boots Socks Aviator Sunglasses Slippers **Novelty Socks Heeled Sandals** Court shoes Cards & Gift Wrap 2 items Flat Sandals Ankle Socks × Click to rename Sneakers Wayfarer & Square Sunglasses

O View instructions

Leave a comment

Finished

Online tool: **OptimalSort** 

prices for cell phone plan option alculate the best cell phone plan for me net connection speed test low to set up my BananaCom email address sfer my home phone number to my new hous ananaCom's contact phone numbe ess to use for BananaCom heli dable games for my cell phone



#### Three types:

#### Open

Participants are asked to organize cards into their own categories.

#### Closed

Participants are asked to sort topics into a set of predefined categories.

## Hybrid

Participants have the ability to sort topics into existing categories or create their own.

## Sample Questions for User Interviews, Focus Groups, Questionnaires, and Workshops

#### For users who search for content

- How do you currently seek/discover information or answers to questions?
- If you use search, what are examples of keywords you use?

#### For users who manage and tag content

- What are the file/document management procedures or workflow procedures?
- How do you organize, categorize, or tag content in these systems?





## **Collaboration in Top-Down Taxonomy Design**

## Focus Groups and Workshops

- Participants respond to others' suggestions and comments
- Shared brainstormed suggestions spawn new ideas
- Seeing other participants contributions in activities contributes to understanding of other views
- Comments from those in other departments foster cross-departmental understanding for future collaboration

#### Interviews

- Interview questions and responses can suggest additional collaboration methods
- Interview subjects can suggest additional colleagues to interview
- Interviews can be done with pairs of people





# Bottom-Up Taxonomy Design

## **Bottom-Up Taxonomy Design Methods**

To ensure that the taxonomy is applicable to relevant content and data sources and meets standards

- Especially focused on the **content**
- Makes more use of technology

#### Methods

- Manual content analysis/survey
- Automated term extraction (corpus analysis)
- Search log analysis
- Keyword tagging analysis
- Legacy vocabularies review
- Term lists from subject matter experts
- LLMs and generative AI

Use when creating a new taxonomy or expanding an existing one.



## **Top-Down Methods**

#### Background Document Review

A review of past or existing documents, content, and items that provides helpful information for the taxonomy design and taxonomy governance efforts.

ENTERPRISE KNOWLEDGE

#### **Content Analysis**

Manually reviewing individual pieces of content (e.g. documents or website pages) to identify patterns of content and possible taxonomies.



#### Automated Term Extraction

The use of text mining entity extraction tools to extract candidate taxonomy terms from a set of documents or web pages.

## Search Log Analysis

Analysis of the spreadsheet generated by the site/enterprise search engine log, indicating what search strings users have been entering into the search box over a period of time.

Bottom-up, but user-focused, rather than content-focused.

#### Procedure

- Sort descending by the number of searches and ignore other data
- Identify and group together similar/synonymous searches
- Designate search strings (including synonyms) with high ranks for inclusion as concepts in the taxonomy
- Choose the preferred concept label and alternative labels

Review search logs periodically for updating the taxonomy, too.

		A	В	С
	1	Search Term	Total Unique Searches	Results Page V
	2	import	104	1.25
	3	extractor	82	1.43
	4	corpus	75	1.16
	5	sparql	73	1.48
	6	api	65	1.12
	7	ontology	64	1.19
	8	excel	56	1.21
	9	export	43	1.09
	10	excel import	40	1.1
	11	workflow	39	1.31
	12	API	38	1.08
	13	graphsearch	38	1.18
	14	snapshot	38	1.16
	15	language	37	1.54
	16	collection	35	1.26
	17	blacklist	34	1.85
	18	suggest	34	1.41
	19	linked data	31	1.23

search log from help.poolparty.biz



## LLMs and Generative AI for Taxonomies



## Challenges with LLMs and Generative AI in Taxonomy Creation

- General LLMs are trained on generic content with a different context than yours and thus different meanings for terms, and different relationships.
- Generative AI (such as ChatGPT) with general LLMs accessing web content generate responses from multiple sources.
  - Taxonomy hierarchies can be incorrectly structured, or the same term can appear in different levels of the hierarchy, when pieced together from divergent sources.
- Generative AI does not recognize copyright and may generate a taxonomy that copies from a copyrighted taxonomy (in entirety or parts).



## LLMs and Generative AI for Taxonomies

#### LLM Uses

- Develop your own LLMs (using Python coding, etc.) to be trained on your own enterprise content for appropriate context and better results
- These LLMs are used for other tasks, such as tagging/classification, and not just for taxonomy creation

#### **Generative AI Uses**

- Create specific, limited queries for sub-tasks of taxonomy development, not for generating an entire, complete taxonomy
  - Suggesting narrower concepts for a concept
  - Suggesting alternative labels for a concept
  - Organizing concepts (e.g auto-extracted terms) into hierarchies
- Generate SPARQL queries to analyze and edit SKOS taxonomies
- Expert taxonomist review of results is always necessary





# Conclusions

## **Conclusions: Value of Design Methods**

## **Top-Down Methods**

- Ensures the taxonomy meets the user needs and expectations.
  - Improving results, so they can do their job better, saving time
  - Improving user satisfaction
- Engages stakeholders, keeping them committed to supporting and improving the taxonomy.

## **Bottom-Up Methods**

- Enuses the taxonomy is relevant for the content.
  - An appropriate number of search results are returned upon selection of a taxonomy concept (not too few or too many).
  - Content is not missed due to lack of taxonomy concepts.





## **Conclusions: Technology for Taxonomy Design**

## **Benefits of Technology**

- Include distributed and remote stakeholders
- Engage stakeholders and users in more ways
- Obtain suggestions via different sources and tools
  - > People respond to different methods differently



#### Collaborative Tools for Taxonomy Design

- Whiteboards for brainstorming and idea sharing: Miro, Mural, Lucidchart
- Collaborative task management software: Trello, Notion, Basecamp
- Spreadsheet for collecting and reviewing terms: Google Sheets, Excel
- Taxonomy management software: PoolParty, Semaphore, Synaptica



## **Conclusions: Collaboration in Taxonomy Design**

## **Benefits of Collaboration**

RPRISE KNOWLEDGE

- Design a taxonomy that is suitable for a larger number of users and uses
- Obtain wider buy-in and long-term support for the taxonomy
- Initiate participation for future maintenance and governance of the taxonomy, which should also be collaborative
- Obtain ideas for additional, expanded uses and implementations of the taxonomy





## **Taxonomy Building Resources**

From Enterprise Knowledge, LLC's website knowledge base:

"The Art of Taxonomy Design"

"The Business Taxonomy Workshop"

"Design Thinking and Taxonomy Design"

"<u>Generative AI for Taxonomy Creation"</u> (by Heather Hedden) From The Accidental Taxonomist blog: "Engaging Others in Taxonomy Building" "Taxonomy Design Research" "Card Sorting and Taxonomies" "Taxonomies and ChatGPT"

From the book <u>The Accidental Taxonomist, 3rd ed.</u>: Chapter 10: "Taxonomy Planning, Design, and Creation"



## **General Taxonomy Resources and Events**

<u>ANSI/NISO Z39.19-2005 (2010) Guidelines for Construction, Format, and</u> <u>Management of Monolingual Controlled Vocabularies</u>

The Accidental Taxonomist Blog

Taxonomy Talk: Taxonomist community on Discord

<u>Bite-Sized Taxonomy Boot Camp London</u> virtual half-day event October 9, 9:00 – 11:45 am EDT

<u>Semantic Data 2024: Taxonomy, Ontology, and Knowledge Graphs</u> October 23, New York, NY

Taxonomy Boot Camp Conference

November 18 – 19, Washington, DC



# Q&A

Thank you for listening. Questions?

Heather Hedden Senior Consultant Enterprise Knowledge, LLC <u>www.enterprise-knowledge.com</u> <u>hhedden@enterprise-knowledge.com</u> <u>www.linkedin.com/in/hedden</u>