



Taxonomy Management 101

Webinar
August 4, 2021

The background of the slide is a light blue-grey color. It features a top-down view of a person's hands using a white tablet. The tablet screen displays a colorful Venn diagram with four overlapping circles in green, yellow, red, and blue. To the right of the tablet, there is a large, complex network diagram with white nodes and connecting lines. A globe icon is also visible within the network. In the bottom right corner, the text "Heather Hedden Data & Knowledge Engineer Semantic Web Company" is displayed in a white, sans-serif font.

Heather Hedden
Data & Knowledge Engineer
Semantic Web Company

About the Speaker



Heather Hedden

Data and Knowledge Engineer
Semantic Web Company

Over 25 years of experience in developing and managing taxonomies, metadata, and other knowledge organization systems for various organizations and applications.

Instructor of self-paced online taxonomy courses.

Prior taxonomy consultant and staff taxonomist.

Author of *The Accidental Taxonomist*.

Semantic Web Company (SWC) and PoolParty



SWC is developer / vendor of **PoolParty Semantic Suite**

Most complete and secure **Semantic Middleware / Semantic AI platform** on the Global Market

W3C standards compliant



ISO 27001:2013 certified

First release in 2009

Current version **8.0**

On-premises or **cloud-based**



Over **200** installations world-wide



Semantic AI:

Fusion of Graphs, NLP, and Machine Learning



Named as Visionary in **Gartner's Magic Quadrant** for Metadata Management Systems 2019, 2020



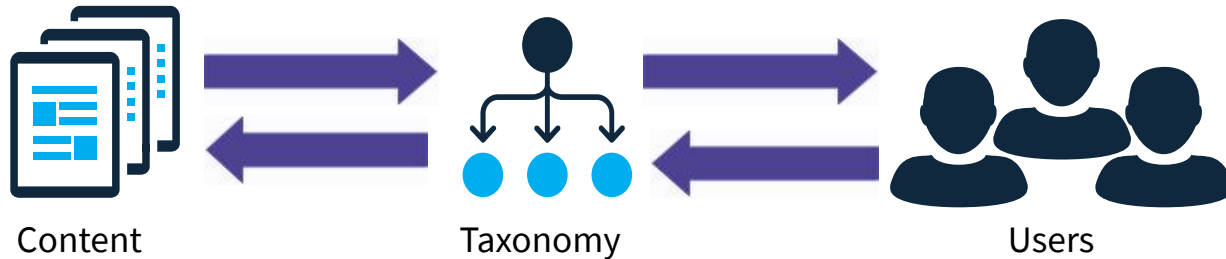
KMWorld listed PoolParty as one of the **Trend-Setting Products** 2015 - 2020 and listed SWC in the **AI 50** list of companies in 2020

- ▶ Why and what are taxonomies?
- ▶ What comprises taxonomy management?
 - ▷ Maintenance
 - ▷ Governance
 - ▷ Tagging management
 - ▷ Review and revision
 - ▷ Extension
- ▶ Who manages taxonomies and how?
- ▶ Taxonomy management tool demo: PoolParty

Why and What are Taxonomies?

Why taxonomies?

- ▶ Concepts/terms are used to tag/index/categorize content to make it 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.

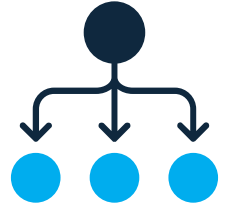


- ▶ Taxonomies usually link to content; ontologies can also link to data.

What is a taxonomy?

Controlled and *organized*

- ▶ A kind of controlled vocabulary or knowledge organization system, based on unambiguous concepts, not just words: **things, not strings**
- ▶ Concepts are arranged in a structure of hierarchies, categories, or facets to organize them.



Why called “taxonomy”?

- ▶ From ancient Greek "taxis," meaning **arrangement** + "nomia," meaning **method**.
- ▶ Originally meant the science and practice of naming and classifying.
- ▶ Aspect of "classification" is still relevant to the definition of taxonomies.

Benefits of taxonomies and other controlled vocabularies

1. Controlled vocabulary

Brings together different wordings (synonyms) for the same concept

- ▶ Helps people search for information by different names



2. Classification and structure

Organizes information into a logical structure

- ▶ Helps people browse or navigate for information
- ▶ Provides context and meaning for concepts for indexing and retrieval



Scope issues: “Taxonomy” sometimes refers to any kind of controlled vocabulary

Why taxonomies?

Purposes and uses of taxonomies and other controlled vocabularies

- ▶ Topic/category browsing
- ▶ Search (matching search strings to concepts)
- ▶ Consistent tagging/indexing
- ▶ Discovery (related concept links, or content sharing the same concepts)
- ▶ Filtering results
- ▶ Sorting results
- ▶ Content management workflow (rights, audience, retention, etc.)
- ▶ Consistent metadata for identification, comparison, analysis
- ▶ Visualization of topics (importance and/or relations)
- ▶ Curated content in feeds or info boxes
- ▶ Automatic linking of relevant topics for personalization or recommendation systems
- ▶ When integrated with ontologies, support for knowledge graphs

Taxonomy Types

Hierarchical taxonomy

Concepts are arranged in a broad-narrower hierarchy

- 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

Faceted taxonomy

Concepts are grouped by aspect

Career Level

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

Function

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

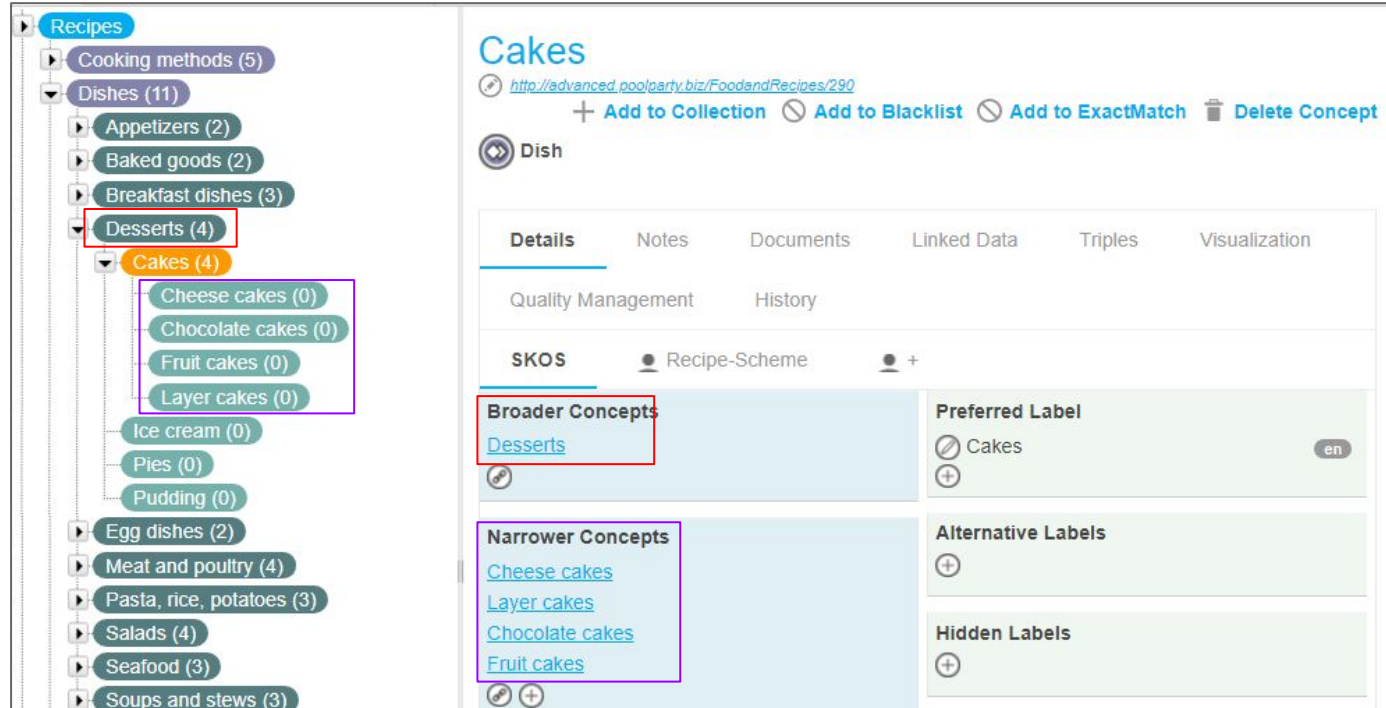
Industry

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

Taxonomy Types

Hierarchical taxonomy

Concepts have broader-concept and/or narrower-concept relationships to other concepts.

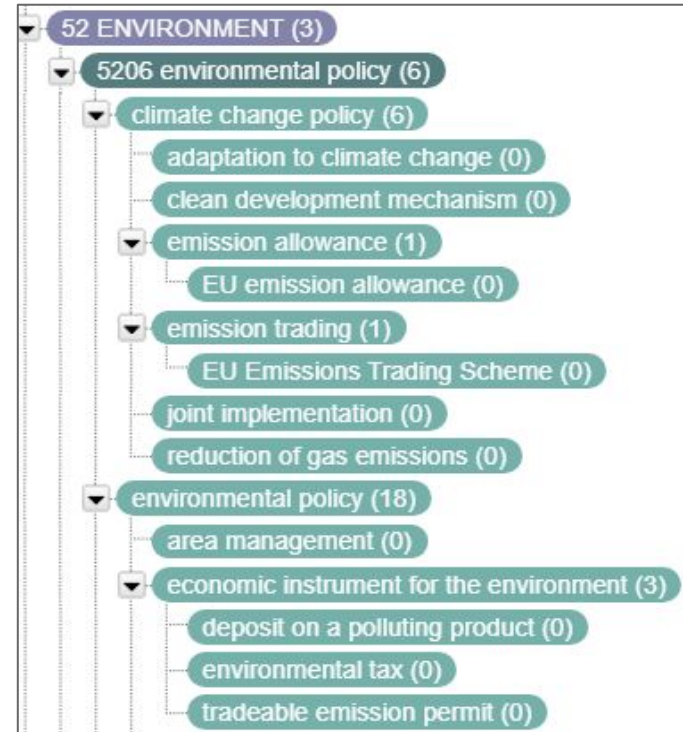


The screenshot displays a taxonomy editor interface. On the left, a hierarchical tree structure is shown under the root 'Recipes'. The tree includes categories like 'Cooking methods (5)', 'Dishes (11)', and 'Desserts (4)'. The 'Desserts (4)' category is highlighted with a red box, and its sub-category 'Cakes (4)' is highlighted with an orange box. Under 'Cakes (4)', four sub-concepts are listed: 'Cheese cakes (0)', 'Chocolate cakes (0)', 'Fruit cakes (0)', and 'Layer cakes (0)', which are enclosed in a purple box.

The right side of the interface shows the detailed view for the 'Cakes' concept. At the top, it displays the URL 'http://advanced_poolparty.biz/FoodandRecipes/290' and several action buttons: '+ Add to Collection', 'Add to Blacklist', 'Add to ExactMatch', and 'Delete Concept'. Below this, the concept is identified as a 'Dish'. A navigation bar includes tabs for 'Details', 'Notes', 'Documents', 'Linked Data', 'Triples', and 'Visualization'. The 'Details' tab is active, showing 'Quality Management' and 'History' sections. The 'SKOS' section is expanded to show 'Recipe-Scheme' with a plus sign. Under 'SKOS', there are three panels: 'Broader Concepts' containing a link to 'Desserts' (highlighted with a red box), 'Narrower Concepts' containing links to 'Cheese cakes', 'Layer cakes', 'Chocolate cakes', and 'Fruit cakes' (enclosed in a purple box), and 'Preferred Label' containing 'Cakes' with a language code 'en' and a plus sign. Below these are 'Alternative Labels' and 'Hidden Labels' sections, each with a plus sign.

Hierarchical taxonomy purposes

- ▶ Serving users who are browsing, exploring, discovering, not searching, to whom the hierarchy is displayed.
- ▶ Providing context to concepts for manual indexers/taggers so that they apply the correct concept.
- ▶ Providing the context of a broader concept and thus meaning to aid in auto-classification.
- ▶ Enabling “recursive”/“rolled up” retrieval results
A concept retrieves what is indexed to it and what is indexed to each on of its narrower terms, all together.
- ▶ Instructing users on appropriate classification.
- ▶ Better organization for controlled vocabulary management

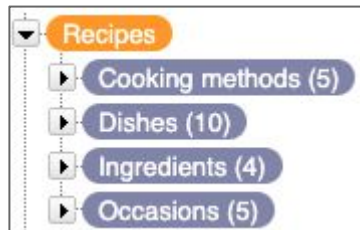


Eurovoc Thesaurus excerpt

Taxonomy Types

Faceted taxonomies

- ▶ Taxonomies organized into aspects, each for a “dimension” of a query
- ▶ Intended for searching with multiple concepts, each from different facets, in combination
- ▶ Each as a refinement, filter, limit by, narrow by
- ▶ Also called “faceted classification” to classify content objects by multiple classification types



Suitable for content of a similar type that shares the same facets.

Examples: all research literature, all internal policies & procedures, all person profiles, all digital media files

The screenshot shows the PoolParty GraphSearch interface. At the top, there is a blue header with the 'GRAPHSEARCH' logo and a green 'Facets' bar. Below this, the 'Document Type' section is expanded, showing a list of search facet values: 'Help Pages - 1199', 'KG Cookbook - 149', 'Gartner Report - 145', and 'Marketing Collaterals - 25'. The 'Epics' section is also expanded, showing a list of search facet values: 'AI strategy - 69', 'explainable AI - 56', 'semantic AI - 34', 'Learning Organization - 32', 'Knowledge management - 31', 'active metadata - 24', 'conversational AI - 18', 'deep text analytics - 7', 'semantic data fabric - 6', and 'GQL - 4'. The 'Industries' section is partially visible at the bottom, showing search facet values: 'service-oriented industry - 122', 'financial services - 77', 'healthcare industry - 53', 'asset-oriented industry - 44', and 'pharmaceutical industry - 29'. The PoolParty logo is visible in the top right corner.

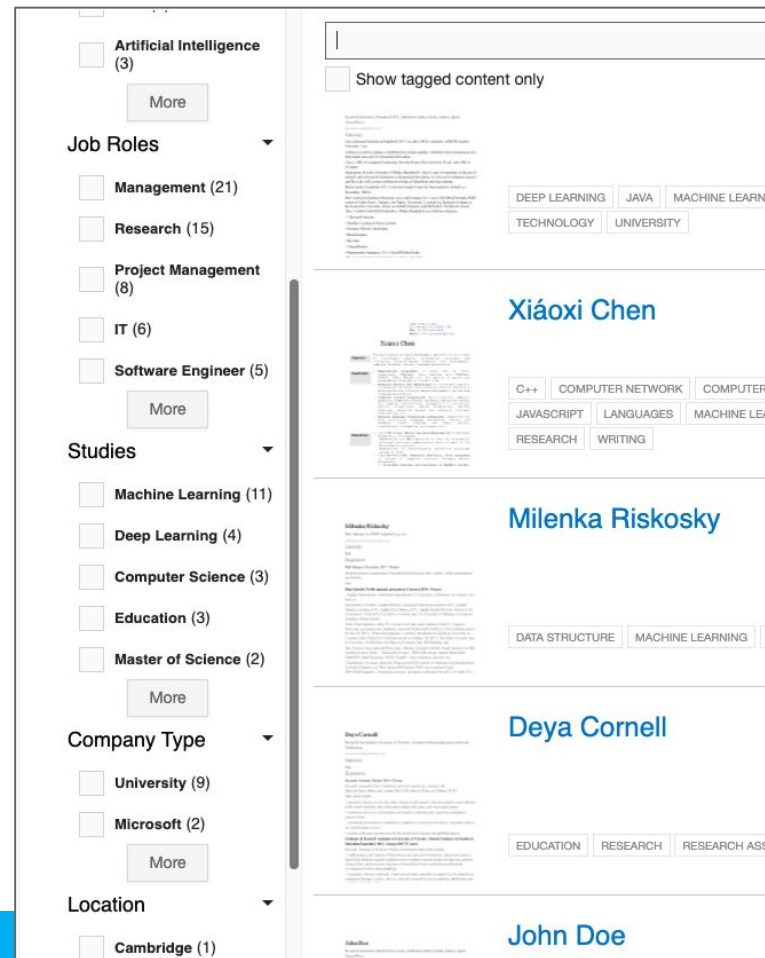
PoolParty
out-of-the-box
faceted search
user interface:
GraphSearch

Taxonomy Types

Faceted taxonomy purposes

- ▶ Ensures comprehensive tagging and comprehensive search/retrieval by multiple different aspects/vocabulary types.
- ▶ Supports filtering search results by different aspects/vocabulary types.
- ▶ Provides guided Boolean “AND” searching upon a combination of concepts in different facets.
- ▶ Allows users to control the search refinement, narrowing or broadening in any manner or order.

Faceted search filters of PoolParty integration in SharePoint.



The screenshot displays a search results page with a left-hand navigation pane containing faceted taxonomy filters. The filters are organized into categories with expandable arrows and 'More' buttons:

- Artificial Intelligence (3)**
- Job Roles**
 - Management (21)
 - Research (15)
 - Project Management (8)
 - IT (6)
 - Software Engineer (5)
- Studies**
 - Machine Learning (11)
 - Deep Learning (4)
 - Computer Science (3)
 - Education (3)
 - Master of Science (2)
- Company Type**
 - University (9)
 - Microsoft (2)
- Location**
 - Cambridge (1)

At the top right, there is a search bar and a checkbox labeled 'Show tagged content only'. Below the filters, the search results are displayed as a list of user profiles, each with a small thumbnail and a list of tags:

- Xiáoxi Chen**: DEEP LEARNING, JAVA, MACHINE LEARNING, TECHNOLOGY, UNIVERSITY
- Milenka Riskosky**: C++, COMPUTER NETWORK, COMPUTER S, JAVASCRIPT, LANGUAGES, MACHINE LEAF, RESEARCH, WRITING
- Deya Cornell**: DATA STRUCTURE, MACHINE LEARNING, M
- John Doe**: EDUCATION, RESEARCH, RESEARCH ASSIS

Taxonomy Types: Thesauri

Thesauri have comprehensive inclusion of:

- Broader concepts
- Narrower concepts
- Related concepts
- Alternative labels
- Scope notes

Thesauri tend not to have only a few extensive hierarchies.

The screenshot shows a taxonomy editor interface. On the left is a tree view of a 'Recipes' taxonomy. The 'Desserts' category is expanded, showing 'Cakes (4)' as a sub-category. Under 'Cakes', there are four sub-concepts: 'Cheese cakes (0)', 'Chocolate cakes (0)', 'Fruit cakes (0)', and 'Layer cakes (0)'. The 'Cakes' concept is selected, and its details are shown in the main panel. The details panel includes tabs for 'Details', 'Notes', 'Documents', 'Linked Data', 'Triples', 'Visualization', 'Quality Management', and 'History'. The 'Details' tab is active, showing the SKOS 'Recipe-Scheme' and several sections: 'Broader Concepts' (with 'Desserts' as a link), 'Narrower Concepts' (with 'Cheese cakes', 'Chocolate cakes', 'Fruit cakes', and 'Layer cakes' as links), 'Related Concepts' (with 'Quick breads' as a link), 'Preferred Label' (with 'Cakes' as the label), 'Alternative Labels' (with 'Tortes' as an alternative label), 'Hidden Labels' (empty), and 'Scope Notes' (with a note: 'A form of sweet food made from flour, sugar, and other ingredients, that is usually baked.').

A light grey background with a network diagram of white circles connected by thin white lines, representing a complex web of relationships or data points.

What Comprises Taxonomy Management?

Taxonomy management includes

- ▶ Ongoing maintenance
- ▶ Taxonomy governance
- ▶ Managing tagging with the taxonomy
- ▶ Periodic review and revision
- ▶ Extension: merging, mapping, translating, integrating, and ontology adding

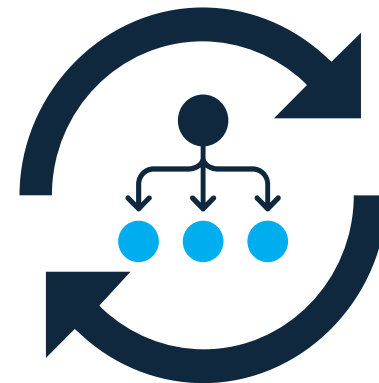
A taxonomy needs to be kept up to date.

Kinds of changes impacting taxonomies

- ▶ Terminology changes
- ▶ New content, bringing up new concepts
- ▶ Content that gets dropped
- ▶ New requirements, users, needs, trends, markets, etc.
- ▶ User feedback suggesting improvements

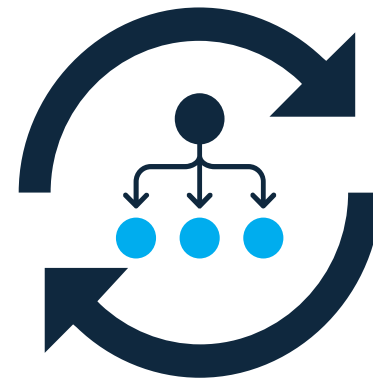
For maintenance, need to review:

- ▶ Newly added content sources
- ▶ Search logs
- ▶ Sections of the taxonomy covering high-change topics
- ▶ Indexing/tagging statistics (manual or automated) to find:
 - ▷ high-use concepts needing further differentiation
 - ▷ low/no-use concepts that should be merged



For maintenance policy and procedures, need to determine:

- ▶ What kinds of changes are routine, which are not
- ▶ What information is needed to determine the changes
- ▶ What group should maintain the taxonomy
- ▶ What role indexers/taggers play in suggesting changes
- ▶ Processes for proposing and resolving changes
 - ▷ Comment-handling, appeals, issue logs, announcements, update schedules, etc.



Taxonomy governance comprises:

- ▶ Maintenance responsibility, roles, processes, procedures
- ▶ Following standards for taxonomies (SKOS and thesaurus standards)
- ▶ Documenting the taxonomy - purpose, type, scope, users, tagging method, history/sources
- ▶ Taxonomy editorial policy/guidelines for maintenance

Taxonomy editorial policy includes:

- ▶ Concept label editorial style (case, abbreviations, acronyms, special characters, diacritics, punctuation, plurals, spelling)
- ▶ Types of concepts included and types excluded
- ▶ Authoritative sources for concepts
- ▶ Guidelines for using alternative labels
- ▶ Hierarchy depth, polyhierarchy policy
- ▶ Scope notes and definitions policy
- ▶ To what extent to follow thesaurus standards (ANSI/NISO Z.39.19 or ISO 25964)



Taxonomy Tagging Management

For manual tagging develop tagging policy, rules, guidelines and documentation

- ▶ Criteria for determining topic or name relevancy for tagging
- ▶ Depth, level of detail
- ▶ Comprehensiveness of aspects (what, who, where, when, how, why, etc.)
- ▶ Required concept types/facets to tag
- ▶ Number of concepts (of each type) to tag

Develop training of those doing tagging.

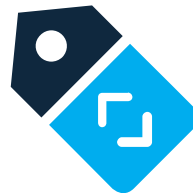
Develop a method of obtaining feedback on the taxonomy from those doing tagging.

For term extraction based auto-tagging

- ▶ Check scoring results and test extraction results.

For auto-classification

- ▶ Train machine-learning based classifiers.



Taxonomy Review and Revisions

Especially for a taxonomy that is inherited, adopted, added, it's needed to review:

Concepts and labels

- ▶ Unambiguous and clear - not deriving its meaning from a broader concept relationship
- ▶ Distinct in meaning - not having two concepts with different preferred labels having almost the same meaning
- ▶ Concise - not too long and wordy
- ▶ Alternative labels sufficiently included, but not excessive

Relationships

- ▶ Hierarchical relationships generally conforming to thesaurus standards, generic-specific (“is a kind of”) or whole-part, not merely grouped subject
- ▶ Related-concept relationship sufficiently included (if used)

Structure

- ▶ Hierarchy levels not too deep, nor too flat
- ▶ Polyhierarchy implemented, as appropriate



Scope and coverage extension

- ▶ By integrating taxonomy: linking, mapping, or merging

Language extension

- ▶ By translating taxonomies or by merging different language versions, creating a single multilingual taxonomy

System extension

- ▶ By plug-ins, APIs, and custom integrations, connecting and synchronizing a centrally managed taxonomy to different front-end applications (CMS, DAM, Intranet, search, etc.)

Semantic extension

- ▶ By adding ontology features of semantic relationships and attributes, based on assigned classes

Linking & Mapping

- ▶ Taxonomies are linked at individual concepts, and the taxonomies are retained as distinct, but can be used in combination, extending each other.
- ▶ **Mapping** is a form of linking for exact or close matches, so that one taxonomy can be used for another, while still retained as distinct.
 - ▷ Used in combination but one taxonomy is the backend, and one is the frontend (not alongside each other).

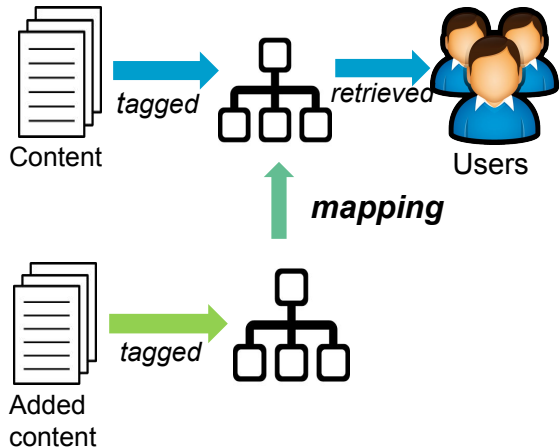


Merging

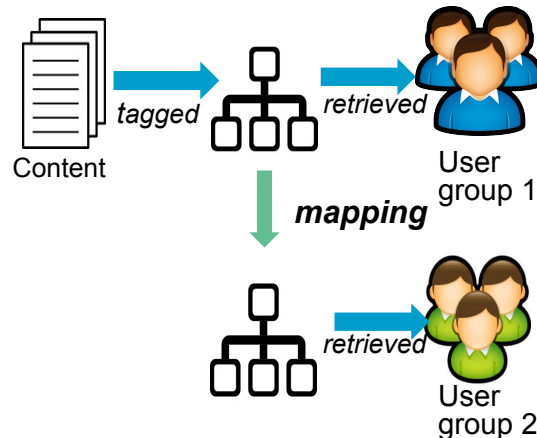
- ▶ Taxonomies are combined permanently, removing duplicates, without any longer retaining them as distinct.
- ▶ First step is to link the taxonomies, then incorporate the unlinked concepts.

Integrating Taxonomies: Mapping Scenarios

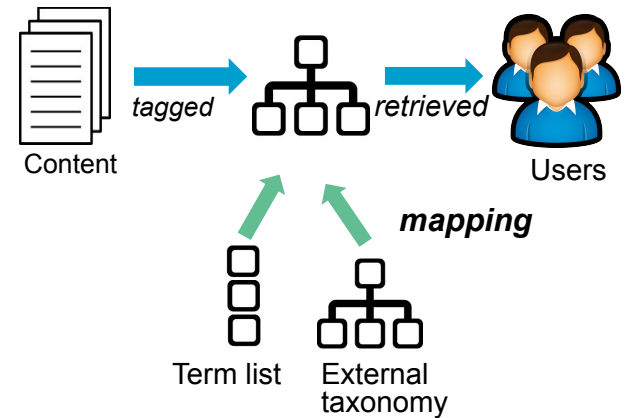
An expanded set of content, tagged with a different taxonomy, will be retrieved by users with their existing taxonomy.



A set of content will be retrieved by different audiences, each accessing their own taxonomy.



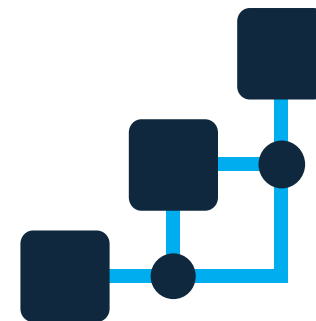
A term list or second taxonomy is mapped to a taxonomy to enrich it.



Adding an Ontology to a Taxonomy

Combination of a taxonomy with an ontology model:

- ▶ An ontology is a model that is applied as a semantic layer to a taxonomy.
- ▶ The ontology comprises classes and subclasses to the extent needed to describe the generic characteristics of the semantic model.
- ▶ The ontology does not include all possible levels of hierarchy, nor any instances.
- ▶ Most of the hierarchy resides in the taxonomy.
- ▶ Additional metadata for characteristics of taxonomy concepts can be managed as ontology attributes.



Adding an Ontology

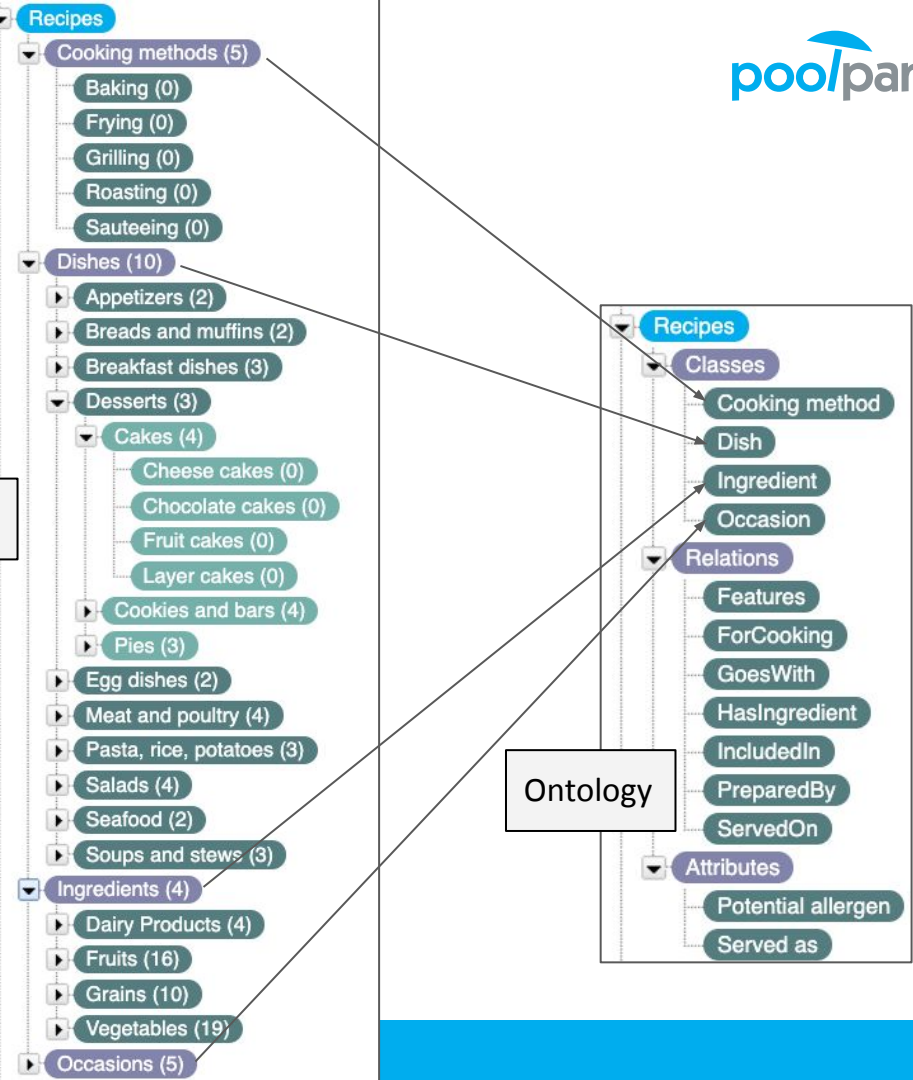
Taxonomy + ontology layer

The ontology tends to be smaller and simpler.

Taxonomy may be based on SKOS, whereas ontology is based on OWL.

Taxonomy

Ontology



Screenshots from PoolParty

Who Manages a Taxonomy and How is a Taxonomy Managed?

Who Manages a Taxonomy?

The creator of a taxonomy is often not its manager.

- ▶ Created as a project by a taxonomy consultant, contractor, or freelancer
- ▶ A legacy taxonomy from a previous project, system, or organization
- ▶ A licensed taxonomy from a third-party publisher

Many organizations lack the resource of a full-time dedicated taxonomist.

- ▶ The taxonomy or taxonomies are not that large.
- ▶ Taxonomy maintenance is not a full-time job.

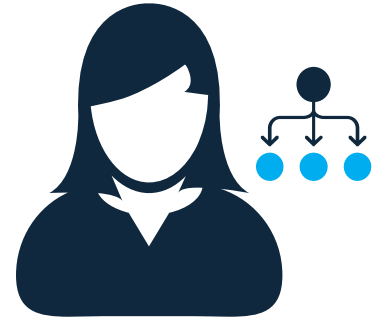
Taxonomy management resides in different departments.

- ▶ Marketing - because the website taxonomy is the biggest
- ▶ Corporate communications - because the taxonomy is part of the Intranet
- ▶ Information technology - because taxonomy management is linked to an IT system
- ▶ Editorial - because the taxonomy is for content publishing

Who Manages a Taxonomy?

If not a “taxonomist” or “taxonomy manager”, there are often others with related skills, who might manage a taxonomy:

- ▶ Content architect
- ▶ Content manager
- ▶ Corporate librarian
- ▶ Data architect
- ▶ Indexing manager
- ▶ Information architect
- ▶ Knowledge architect
- ▶ Knowledge manager
- ▶ Product manager (for information products)
- ▶ Search analyst
- ▶ Solutions architect
- ▶ UX/UI designer or architect



Who Manages a Taxonomy?

Taxonomy management should not be done in isolation.

Collaboration is needed with:

- ▶ Product owners
- ▶ Content owners or creators
- ▶ UX designers/information architects
- ▶ Search specialists
- ▶ IT systems developers

By means of:

- ▶ Standing meetings, check-in meetings
- ▶ Review processes
- ▶ Feedback meetings/surveys/interviews



How is a Taxonomy Managed?

How to manage a taxonomy

- ▶ **Scope:** Determine the taxonomies to manage in different systems
- ▶ **People:** Determine roles and responsibilities and determine stakeholders to include
- ▶ **Input:** Establish a feedback mechanism for tagging success and for the front-end user experience success
- ▶ **Technology:** Manage the taxonomies centrally, not in individual systems. Use a dedicate taxonomy management system that integrates with other systems.

A light grey background with a network diagram consisting of various sized circles connected by thin lines, representing a complex web of relationships.

Taxonomy Management System PoolParty

Why PoolParty for taxonomy management?

- ▶ Central management of a taxonomy for multiple uses/integrations (with APIs)
- ▶ Supports all aspects of taxonomy management
- ▶ Easy-to-use user interface
- ▶ Manages different types of taxonomies, thesauri, name authorities, term lists, glossaries, ontologies, etc. with all of their features
- ▶ Analytics, metadata, and visualizations of the taxonomy
- ▶ Linking/mapping of multiple taxonomies and vocabularies
- ▶ Semantic Web standards of SKOS, RDF, OWL, and validates data
- ▶ ANSI/NISO and ISO thesaurus standards
- ▶ Multiple users with different levels of access and approval workflows
- ▶ Multi-lingual taxonomies
- ▶ NLP-based term extraction and auto-tagging capabilities; ML-based auto-classification
- ▶ Knowledge graph creation and editing, with embedded graph database and option KG data management and editing features.

Taxonomy Management System: PoolParty

PROJECT CORPORA TOOLS ADVANCED en Search Thesaurus Concepts

Workflow Dashboard

Appetizers

<https://elysium.poolparty.biz/Recipes/107>

Dish

+ Add to Collection Add to Blacklist Add to ExactMatch Delete Concept

Details Notes Documents Linked Data Triples Visualization Quality Management History

SKOS Recipe-Scheme

Broader Concepts

Narrower Concepts

- [Bruschetta](#)
- [Dips](#)
- [Quesadillas](#)
- [Stuffed mushrooms](#)

Related Concepts

Top Concept of Concept Schemes

- [Dishes](#)

Preferred Label

- Appetizers

Alternative Labels

- Hors d'oeuvres

Hidden Labels

- Appetisers
- Horderves

Scope Notes

- Dishes usually served as appetizers

Definitions

- A small dish of food or a drink taken before a meal or the main course of a meal to stimulate one's appetite

PoolParty Demo

Recipes

- Cooking methods (4)
 - Baking (0)
 - Frying (0)
 - Roasting (0)
 - Sauteeing (0)
- Dishes (10)
 - Appetizers (4)
 - Bruschetta (0)
 - Dips (2)
 - Guacamole (0)
 - Hummus (0)
 - Quesadillas (0)
 - Stuffed mushrooms (0)
 - Breads and muffins (2)
 - Breakfast dishes (3)
 - Desserts (3)
 - Egg dishes (2)
 - Meat and poultry (4)
 - Pasta, rice, potatoes (3)
 - Salads (4)
 - Seafood (3)
 - Soups and stews (2)
- Ingredients (3)
 - Dairy Products (5)
 - Grains (7)
 - Vegetables (20)
- Occasions (4)
 - Christmas (0)
 - Grilling (0)
 - Parties (0)

- ▶ Other PoolParty webinar recordings and white papers
<https://www.poolparty.biz/resource-library>
- ▶ The Accidental Taxonomist Blog
<http://accidental-taxonomist.blogspot.com>
- ▶ ANSI/NISO Z39.19-2005 (2010) *Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies*.
www.niso.org/publications/ansiniso-z3919-2005-r2010
- ▶ SEMANTiCS conference, September 6-9, 2021, Amsterdam & online (hybrid)
<https://www.taxonomybootcamp.com>
“Tutorial: Knowledge Engineering of Taxonomies, Thesauri, and Ontologies”
- ▶ Taxonomy Boot Camp, November 15-16, 2021, Washington DC
<https://www.taxonomybootcamp.com>



Heather Hedden

Data and Knowledge Engineer
Semantic Web Company Inc.
One Boston Place, Suite 2600
Boston, MA 02108
USA

+1 857-400-0183

heather.hedden@semantic-web.com

www.linkedin.com/in/hedden

Semantic Web Company www.semantic-web.com

PoolParty software www.poolparty.biz