Extending Taxonomies to Ontologies

Heather Hedden

Sometimes the words "taxonomy" and "ontology" are used interchangeably, and while they are closely related, they are not the same thing. They are both considered kinds of knowledge organization systems to support information and knowledge management. Yet there is often a lack of agreement on their definitions, although published standards help define them both. Rather than debating definitions, what is of greater importance is what a taxonomy or ontology enables you to do.

Benefits of Taxonomies and Ontologies

Taxonomies (hierarchical or faceted structured controlled vocabularies of concepts) primarily enhance search and retrieval of content, but they have related benefits. Taxonomy uses and benefits include:

- Tagging: to index content consistently so that retrieval is comprehensive and accurate
- Normalization: to bring together different names, localizations, and languages for concepts
- Standard search: to enable users to find content about something (whereby the user's search string matches taxonomy concepts)
- Topic browse: to enable users to explore subjects arranged in a hierarchy and then get content on the selected subject
- Faceted (filtering/refining) search: to enable users to find content that matches a combination of basic criteria
- Discovery: to enable users find additional, related content tagged with the same concepts; to explore broader, narrower, and (sometimes) related taxonomy topics

- Content curation: to create feeds or alerts based on pre-set search terms
- Metadata management: to support identification, comparison, analysis, etc., in addition to content retrieval

Ontologies (semantic models comprising the types/classes, semantic relationships, and attributes of entities) were originally for describing a domain while also supporting inference for learning more about the domain. However, when entities from a taxonomy are combined with an ontology, benefits and capabilities include:

- Modeling complex interrelationships (e.g. in product approval or supply chain processes) while also connecting to content
- Executing complex multi-part search queries
- Exploring explicit relationships between concepts, not just broader, narrower, or related
- Searching across datasets, not just searching for content
- Searching on more specific criteria that vary based on category (class)
- Visualizing concepts and semantic relationships
- Reasoning based on inferences
- Creating knowledge graphs (incorporating instance data), upon which additional knowledge applications can be built

"Content" refers to files, documents, images, intranet pages, spreadsheets, etc. "Data" refers to such things as the information within database records and the cells within tables or spreadsheets. Sometimes people are looking for content, sometimes they are looking for data, and sometimes they are looking for both. Taxonomies focus on connecting users to content, and ontologies focus on data, so a combination of taxonomies and ontologies can connect users to

both content and data, in addition to connecting the content and data together.

Taxonomies and Ontologies Combined

Taxonomies and ontologies have different origins (library/information science vs. computer/data science), and thus usually different experts, but these two knowledge organization systems have converged greatly in the past decade. There are two primary reasons for this convergence:
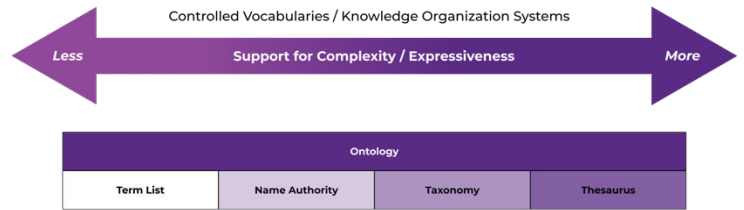
- The adoption of shared Semantic Web (World Wide Web Consortium) standards, whereby both taxonomies and ontologies are built on the same data model, RDF (Resource Description Framework), and other models and standards based on RDF. Thus they can be built in the same tools and connect to each other seamlessly.
- The increased business needs to manage and extract knowledge from growing volumes of content and data together in sophisticated ways as well as the growing demand for data and information, not just for documents and pages.

As mentioned above, there are different definitions for ontologies, and a leading difference concerns whether individual entities are included within the scope of "ontology." An ontology is either:
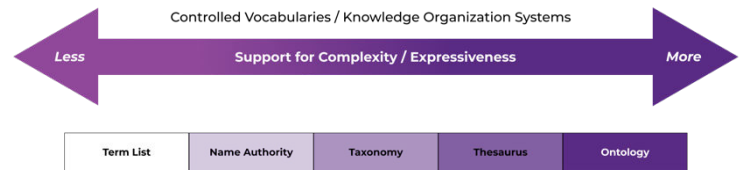
1. A model of a knowledge domain, comprising classes, semantic relationships, and attributes (along with prescribed rules or constraints on each of these components, etc.), or
2. A model of a knowledge domain, comprising classes, semantic relationships, and attributes, *plus* all the individual members of the classes, which are described in controlled vocabularies, including taxonomies.

The following pair of diagrams listing different controlled vocabulary and knowledge organization systems illustrate the views of these two different definitions of ontologies.

1. Ontology as a model of a knowledge domain that serves as a semantic *layer* connected to various controlled vocabularies:

Controlled Vocabularies / Knowledge Organization Systems

Less — Support for Complexity / Expressiveness — More

Ontology			
Term List	Name Authority	Taxonomy	Thesaurus

2. Ontology as a model of a knowledge domain that serves as a semantic *layer* connected to various controlled vocabularies:

Controlled Vocabularies / Knowledge Organization Systems

Less — Support for Complexity / Expressiveness — More

Term List	Name Authority	Taxonomy	Thesaurus	Ontology

Depending on how you define ontology, above, a taxonomy can then either

1. be *enhanced* to include an ontology as an additional semantic layer (definition #1), or
2. be used as an important *component* of an ontology (definition #2)

Ontologies alone may have taxonomic *features* of deep hierarchies of classes and subclasses, but without a taxonomy or thesaurus built on the SKOS (Simple Knowledge Organization System) data model, the full range of functionality of alternative labels, labels in other languages, multiple definitions and types of notes, etc. are not supported. Taxonomies provide a linguistic aspect that ontologies alone lack.

Ontologies alone would support modeling, exploring, and visualizing entities and their relationships, which may be based on their properties. Ontologies may also support inference reasoning. However, functions involving semantic search, which brings together synonyms and disambiguating homonyms, etc. require taxonomies, thesauri, or other controlled vocabularies.

Creating an Ontology Based on Taxonomies

Regardless of which of the two definitions of ontology you prefer, if you already have a

taxonomy, which is often the case, you can extend it to become or or add an ontology and then reap the additional benefits of the combined knowledge organization system. If you have multiple taxonomies and other controlled vocabularies, an ontology can link them together.

Whether you are building a taxonomy, ontology, knowledge graph, or a broader digital transformation for knowledge management, there should be a combination of top down and bottom up approaches to the process. The top-down methods focus on obtaining input from stakeholders, whereas the bottom-up methods focus on analysis of content and data.

The basic approach to building an ontology, especially a business or enterprise ontology, is to identify groups of things (or "business objects"), which become classes in an ontology, identify relationships between pairs of classes, and identify important characteristics (or attributes) of members of a class. The top-down approach to this task involves interviewing stakeholders and conducting brainstorming sessions and focus group sessions to identify these classes, relationships, and attributes. The bottom-up approach to ontology creation often involves looking at spreadsheets and tables of critical data pertaining to different business objects.

A quicker bottom-up approach to creating ontologies is to look at the taxonomies and controlled vocabularies you already have. Each taxonomy hierarchy, controlled vocabulary, term set, facet, or what is designated as a "concept scheme" in the SKOS model can be considered to be a class in an ontology. Additional classes or subclasses might get added, and some term lists might not be needed in an ontology, but often concept schemes can serve as the basis of classes, one-to-one.

Facets in a faceted taxonomy enable browsing or limiting searches for content items by certain aspects. However, content needs to be limited to that of a similar kind that shares the same facets, such as all product pages, all reports, all employee profiles, or all media files. If we can convert the facets to ontology classes, create new semantic relationships between them, and tag all content, a search application is no longer limited to a certain kind of content or asset. Rather, conditional queries in the same application/user interface can be targeted at any kind of content.

Example: Converting Facets to Classes to Build an Ontology

Consider an example for an organization's internal knowledge base. There may exist multiple repositories of content and data, each with its own faceted taxonomy and its own user interface.

- **Reports** could be searched using a Reports faceted taxonomy, which has the facets Report Type, Subject, Author Name, and Division.
- **Employees** as experts could be searched using a People faceted taxonomy, which has the facets Name, Job Title, Location, Division, Skills, and Subject Expertise.
- **Media files** could be searched using a Digital Asset Management faceted taxonomy, which has the facets Subject, Location, Event, Person Depicted, and Creator.

We could create classes to reflect the aggregation of all of these facets.
- Division
- Employee name (which also includes report authors and media asset creators)
- Event
- File type (with subclasses for Document type and Asset type)
- Job role (including titles)
- Location
- Skill
- Subject (including expertise areas)

Then we could consider the relationships or links between the classes, and create verb-based semantic relationships. Any class that is a target/object of a relationship can be a target of a search query. The following are just some examples, but not a complete list with all reciprocal relationships.

Employee *knows* Subject
Employee *created* File Type
Employee *possesses* Skill
Employee *basedIn* Location
Employee *belongsTo* Division

File Type *hasTopic* Subject
File Type *createdBy* Employee
File Type *belongsTo* Division

Subject *knownBy* Employee
Subject *topicOf* Division
Subject *topicOf* File Type
Subject *topicOf* Event

Event *basedIn* Location
Event *belongsTo* Division
Event *hasTopic* Subject

Finally, you should consider what additional data is of importance for the entities in each class, such as contact information for Employees and dates of publication for files and for the occurrence of Events. These would normally not exist in a taxonomy, but should be added to the ontology to support the exploration of more kinds of data.

Conclusions

Combining a taxonomy with an ontology provides many benefits and capabilities which a taxonomy alone or an ontology alone (as merely a semantic model) cannot provide.

Building an ontology based on one or more existing taxonomies is an efficient and very suitable method of bottom-up development. The existing taxonomies and controlled vocabularies provide a basis for knowledge modeling. Furthermore, by leveraging an existing taxonomy that has already been tagged to content, certain benefits of the ontology will already be in place.

Managing the taxonomy plus the ontology as a semantic layer also has benefits.
A taxonomy plus ontology is more flexible and adaptable than a single large ontology, since the taxonomy changes more frequently than does the ontology. Also, more taxonomies and controlled vocabularies can easily be added in the future. There are also several software options for combined taxonomy-ontology creation and management. These applications are based on RDF, including SKOS for taxonomies and RDF-S and OWL for ontologies. This facilitates the technical aspects of extending a taxonomy to become an ontology.

Although extending taxonomies to become ontologies is easier than creating ontologies from scratch, it still requires ontology design expertise. For assistance in extending your taxonomies into an ontology, contact us to get started.

Enterprise Knowledge (EK) is a services firm that integrates Knowledge Management, Information Management, Information Technology, and Agile Approaches to deliver comprehensive solutions. Our mission is to form true partnerships with our clients, listening and collaborating to create tailored, practical, and results-oriented solutions that enable them to thrive and adapt to changing needs.

Our core services include strategy, design, and development of Knowledge and Information Management systems, with proven approaches for Taxonomy Design, Project Strategy and Road Mapping, Brand and Content Strategy, Change Management and Communication, and Agile Transformation and Facilitation. At the heart of these services, we always focus on working alongside our clients to understand their needs, ensuring we can provide practical and achievable solutions on an iterative, ongoing basis.

info@enterprise-knowledge.com | 571-403-1109 | @EKConsulting