

Taxonomy: Sort, Search and Simplify Your Digital Assets

optimityadvisors.com



If you have been involved in creating, marketing or distributing content for a consumer-facing organization, you have probably heard that taxonomy is a great way to improve content storage and distribution. The concept has generated a lot of buzz, but remains a mystery to those who are not information professionals. Many leaders in business have come to understand that taxonomy is good for their digital information, but do not necessarily understand exactly how it positively affects the structure, flow, storage and overall findability of their content.

Taxonomy does not have to be a mystery. With some basic knowledge of what taxonomies are, how they relate to assets, content, and brands, and the steps involved in successfully implementing taxonomies, this useful tool can be utilized in virtually any organization.

TAXONOMY IN INFORMATION MANAGEMENT

Optimity Advisors defines taxonomy as the classification of information into groups or classes that share similar characteristics. The most classic and recognizable example of a taxonomic structure is that of the animal kingdom, which breaks down the entire kingdom of Animalia into its respective groups, getting progressively more specific as it continues down the hierarchy from kingdom and phylum to genus and species. Use of taxonomies was largely relegated to biology until the computer age ushered in a new period in which information was as bountiful as life itself, and required a similarly structured method of organization.

Taxonomies became popular in business around the time the internet became a ubiquitous way to interact with consumers. As software and websites became more sophisticated and information, assets and data proliferated, the need to organize content became imperative. Enter the taxonomy. Since their

inception for business use, taxonomies have been implemented to organize information for e-commerce websites, intranets, document management, knowledge management, search, digital asset management and much more. The more information that has to be sifted through, the more imperative that it is organized using taxonomy and metadata.

TAXONOMY BENEFITS

Creating a hierarchical classification system around a company's knowledge, assets, products, content, brand, etc., has countless benefits for the organization. The pain points that lead companies to seek control over their content can vary, but ultimately the savings in time and reduction in compliance risk—as well as the potential to easily discover assets years after they were created—are what drive companies to create a unified taxonomy.

Support system integrations and increase data quality

Companies have recently begun to see the folly of their ways in setting up numerous information systems in silos of business group-specific needs. Eventually they notice that these different business groups with their unique ways of describing things have created data that has the same intent, but may have a different **lexicon**.¹ A taxonomy will normalize these vocabularies so that data can transfer across systems seamlessly, and data quality is maintained at the highest standard.

Enable discovery and improve findability

Taxonomies enable discovery by consistently describing and correctly spelling content in information systems, and avoiding issues with findability due to **transcontinental spelling differences**.² Navigation can more accurately drive users to the content they need when it is properly **tagged**.³



Taxonomy Benefits (continued)

Develop a consistent voice and brand

Consistency of voice and brand across channels both locally and globally is a benefit whose value cannot be overestimated. By adding a global taxonomy to an information system, an entire organization can have access to the latest in brand evolution without missing a beat. This not only creates a common language across traditionally siloed departments, but the latest media communications will use the terminology that accurately reflects the brand.

Reduce duplicative efforts and wasted time

Improved findability of assets enables users to easily access existing content, improving their productivity by eliminating the back and forth of email, phone calls and searches of servers and hard drives.

Facilitate process automation and lifecycle management

Generally, taxonomies are built in tandem with a **metadata model**⁴ and accompanied by a controlled vocabulary. Combined, these elements create the standards that enable automation. They capture the necessary information supporting automated data flows, approval and comment receipt during work-in-progress phases, and timely decisions around content disposition and archiving.

Provide a basis for content reuse

One of the most exciting benefits of a taxonomy is the potential for reusing assets that may have otherwise been lost to time. Too often, historical information about programs, campaigns, organizational changes, etc. disappears when the last staff member with knowledge of it departs the organization. But this can be prevented when content is well-documented and tagged, including updating information as changes occur and brands evolve. Up-to-date taxonomies evolve as a business does, and changes can be tracked alongside content so past revisions may still be found, used and reused.

Safeguard regulatory compliance

A key benefit of classifying business content is the ability to account for regulatory compliance and improve standardization around these processes. Taxonomies support standards of practice (SOPs) by, for example, differentiating assets that are to be used in different markets or regions, keeping embargoed or restricted assets inside the firewall, and unifying processes into standard models that are less subject to human error.

Improved user adoption

User adoption is an ever-present threat to new information system implementations. Low user adoption can lead to unexpected costs, whether from enacting stringent change management, or having to scrap a system and start all over. But a good taxonomy that is built around user needs—with an eye toward the user experience—can improve adoption in several ways. For example, involving users in a new system implementation early makes them feel that they have had a voice in its development, rather than feeling that it has been foisted upon them. Additionally, because the end users are the ones informing the taxonomy, they design a system that will truly suit their needs, which naturally encourages use.

Search enablement

A taxonomy enables search across the enterprise, but that is just the beginning. In time, complex, large-scale taxonomies can develop **spanning topics** from all over the organization. In fact, these large taxonomies can morph into **ontologies** and work together with sophisticated algorithms to create natural-language processing and near-artificial intelligence capabilities that take information management to the next level of automation and big data processing.

¹ The vocabulary of a person, language, or branch of knowledge.

² British and American spelling often differs. British English kept many European spellings as new words were adopted into the English language, and American English spellings instead reflect the way those words sound aloud (e.g. authorize / authorise).

³ By “tagging” content, you are describing it with keywords. This is to make it easily search-able.

⁴ A metadata model is a representation, or blueprint, of how metadata is stored in your digital assets.



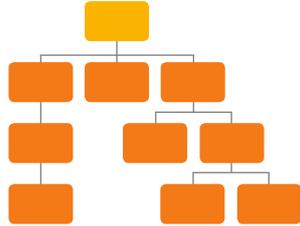
TAXONOMY BASICS

The classic taxonomy is a hierarchy that starts with a single top node and expands into a tree-like structure, becoming increasingly specific as the nodes branch out. The structure has evolved as technological advances have allowed information description to expand into polyhierarchies, faceted taxonomies and ontologies.

Types of Taxonomies

Hierarchical Taxonomy

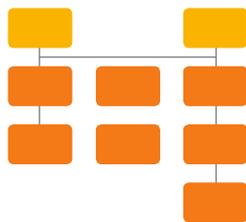
Hierarchical taxonomies are represented as a “tree” data structure in a database application. The tree data



structure consists of nodes and terms. Nodes are top level terms that will have more specific classifications branching out below them. These classic, highly usable data structures are implemented across information systems in a “hidden in plain sight” kind of way. Think about left-hand and even top-level navigation of any website, and you are thinking of a classic hierarchical taxonomy. This is the basic information organization structure, to which you might add synonyms, relationships, links, definitions and other levels of information to increase its complexity.

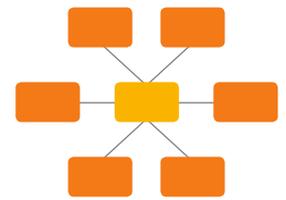
Polyhierarchies

Polyhierarchies are similar to hierarchies, with an added layer of complexity. With a strictly hierarchical taxonomy, a node can have only one parent. Strictly classifying ideas into singular concept classes prevents them from getting mixed up across the structure. An example of this can be seen in the topic of Health Education. Strictly speaking, this is a form of education and would only belong under the Education node. However, when creating taxonomies for navigation, it can be useful to create a polyhierarchy, in which Health Education also exists under the Health node, so that users who are navigating the system can find this topic under either of those major categories. A word of caution, however: polyhierarchies should be used sparingly, and only in the case of a simple navigational taxonomy. Overuse of a polyhierarchical structure in a larger, more complex taxonomy can lead to confusion in classification.



Faceted Taxonomy

Unlike the tree structure of a hierarchical taxonomy, a faceted taxonomy is represented as a “star,” with each node in the star structure linked to a center focus. Also, any node can be its own star, acting as the center focus of its own small universe of topics. The trick is that any node can be linked to other nodes in the universe, including nodes of stars at different levels. A faceted taxonomy appears simple at first, but it quickly becomes complex.



Basic Components of Taxonomy

Primary Term – The agreed-upon term that stands as the official terminology for a given product, concept, brand, location, etc. It is used as the front-facing tag for a given asset.

Synonym – The secondary term or terms that facilitate discovery of assets even when the preferred terminology is not known.

Related Term – Any term that might be related to another term without being in the same hierarchy. To use the earlier example, Health Education exists in the Education branch of the taxonomy, but it is related to Health.

Node – The terms in taxonomy are known as nodes. The terms that represent the first layer of nodes—i.e., the broadest terms in the taxonomy—are known as top nodes.

Scope Notes – These notes clarify any possible uncertainty as to the scope of a term’s definition and how it should be used in indexing. These will often be found in the preferred and non-preferred terms to define how they relate to each other as synonyms.

Authority File – Allows for setting precedents or rules around the definition, usage or changeability of a term.

Broader Term / Parent Term – A less specific term in a hierarchy. Usually has two or more child terms. Can have a parent term of its own.

Narrower Term / Child Term – The more specific term in a hierarchy. Can have child terms of its own.



BUILDING A TAXONOMY STRATEGY

Many companies fail to do the extensive groundwork that is required to develop a solid taxonomy. It is not enough to want a taxonomy for your information solely because it may help improve sales. A solid goal for the project and understanding of the information landscape are vital to the success of any taxonomy initiative.

Define the goals of the project

An imperative first step to developing a strong taxonomy strategy is to understand what the problem is that you are trying to solve. Are there bottlenecks in the asset development workflow? Are users spending too much time searching for assets that are spread out across systems and drives? Are there issues with maintaining a consistent brand voice? Are consumers struggling to navigate around your content?

Learn about the behavior of system users

Whether developing a taxonomy for an anticipated new system, or revamping the taxonomy for an existing system, knowing how users will use, interact with, ignore, update, etc., the different elements will have an influence in how the taxonomy develops. For instance, user search behavior should influence the design of a search taxonomy. Navigation and storage habits should influence a navigational taxonomy. And the overall information needs that the system satisfies should influence the vocabulary used to define its contents.

The simplest and most effective way to understand how users interact with systems is to watch them in action. This includes everything from asset creation and storing to seeking and sharing. These various aspects of the asset lifecycle will give a rounded understanding of the language that people use to describe assets, as well as how they prefer to search, navigate and otherwise interact with the system. Another way to see how people use search is to review existing search logs, either of other systems if there is no system in place yet, or of the existing system if it is an upgrade project.

Consider the basic tenets of controlled vocabularies

There are many considerations around the development of taxonomies, and these are outlined by the National Information Standards Organization (NISO) “Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies.”⁵ **Of course, many good books have been written on the development of taxonomies,**⁶ but this is one of the most extensive and definitive guides and references for the task.

The NISO guide outlines high-level and granular details about how a controlled vocabulary, including a hierarchical taxonomy, should be structured, right down to which words should be plural and which should be singular. But some key elements to keep in mind before embarking on a taxonomy development project include:

- Make sure there is content at every level. That is to say, do not build nodes for future content. When building out the vocabulary, it can be tempting to create a node for forthcoming topics, but this practice should be avoided.
- Ideally, build the taxonomy out broadly, avoiding going more than three layers deep. It can be tempting to develop a taxonomy that goes several layers deep into the tree structure. But this requires users to click and click and click to find what they are looking for. People prefer to scan a page for information, and they expect to find what they seek within a click or two.
- It is also good practice to try to balance breadth and depth in the taxonomy. A taxonomy that has 35 top nodes and only two categories per node has probably missed an opportunity to group similar concepts into classes. Alternatively, a taxonomy that has four top nodes but goes eight layers deep in each node may have mixed concepts together that should be differentiated.
- Taxonomies have branches, and branches should have “leaves.” The “leaves” represent the categories that reside within each branch of the taxonomy. Just as no node should be without content, each branch of the taxonomy should have at least two categories. This is because there should be sufficient content in each category to warrant its existence. The practice prevents an explosion of content categories that could logically belong together, even if they have minor differences that make them seem unique. Refrain from getting bogged down in the minutiae of differences between concepts, and remember that a good taxonomy does not lead users down a rabbit hole to find nothing at the other end.

⁵ <https://www.niso.org/publications/ansiniso-z3919-2005-r2010>

⁶ Inform, Transform, & Outperform: Digital Content Strategies to Optimize Your Business for Growth
ANSI/NISO Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies
Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness
Building Enterprise Taxonomies
Typologies and Taxonomies: An Introduction to Classification Techniques



TAXONOMY GOVERNANCE

Information systems cannot just be set up, implemented and left to run on their own—including taxonomies. Unfortunately, many assume just that. But because taxonomy is based in language, it is subject to language's changeability. Just think about how much our vocabulary changes year after year. New words are added to the dictionary, organizations launch campaigns with new buzzwords, digital slang migrates to verbal communication—taxonomy initiatives fall prey to the same fluctuation.

Therefore, those inevitable shifts must be considered when implementing a taxonomical system. Without careful planning and upkeep, the taxonomy across various systems can devolve, with consequences that include unchecked additions on new nodes, unarchived retired concepts and unimplemented term changes. To avoid these issues, put an information professional in place who understands the basic tenets of information organization and cares about your system's longevity.

CONCLUSION

Optimity Advisors understands the significant impact a taxonomy can have on organizing information for discovery, findability, storage and lifecycle management. Regardless of whether you're still learning the basics of taxonomy, or you're ready to construct a strategy, Optimity Advisors can guide you on the journey to your future state. Our information professionals utilize the basic tenets of information organization, alongside international standards for taxonomic development and developing a deep understanding of an organization's specific lexicographical needs, to help our clients sort and standardize their digital house.

We know that taxonomy is more than a buzzword—it is a powerful organizational tool that, when wielded correctly, can lighten workloads and streamline processes. The implementation of a taxonomy is not only a smart move, but in this digital age, it has become a necessary one.



Mindy Carner is one of Optimity's premier information science subject matter experts. She has a Master's Degree in Library and Information Science, and nearly ten years of experience creating, deploying, managing and governing metadata and taxonomy strategies for national and international organizations. As a Senior Manager, Mindy supports clients on a variety of information system deployments by leading information discovery, content assessments, metadata and taxonomy strategy, vendor selection, implementation management, deployment and change management strategy and governance planning.



THANK YOU

Thank you for your valuable time. For further information, please contact:

Mindy Carner, Senior Manager

Direct: (917) 242-0420

Mindy.Carner@OptimimityAdvisors.com



optimimity®
A D V I S O R S

INFORM | TRANSFORM | OUTPERFORM

Washington, D.C. • Berlin • Brussels • London • Los Angeles • Minneapolis • New England • New York • Zurich

Optimimity Advisors helps you design and implement new strategies, products, services, processes, business models and applications that transform your business. Together, we create better experiences — experiences that drive measurable value in a hyper-connected world.