



Date : 24/07/2008

## Taxonomy Development for Knowledge Management

**Mary Whittaker**

Librarian

Boeing Library Services

The Boeing Company

PO Box 3707 M/C 62-LC

Seattle WA 98124

+1-425-306-2086

+1-425-965-0119

[mary.s.whittaker@boeing.com](mailto:mary.s.whittaker@boeing.com)

**Kathryn Breininger**

Manager

Boeing Reports Management Services

The Boeing Company

PO Box 3707 M/C 62-LC

Seattle WA 98124

+1-425-965-0242

+1-425-512-4281

[kathryn.r.breininger@boeing.com](mailto:kathryn.r.breininger@boeing.com)

**Meeting:**

**138 Knowledge Management**

**Simultaneous Interpretation:**

English, Arabic, Chinese, French, German, Russian and Spanish

---

WORLD LIBRARY AND INFORMATION CONGRESS: 74TH IFLA GENERAL CONFERENCE AND COUNCIL

10-14 August 2008, Québec, Canada

<http://www.ifla.org/IV/ifla74/index.htm>

---

### **Abstract**

*As librarians and information professionals, we are faced with solving a growing information overload problem. We strive to connect end users with the information they need. How can we better connect searchers to the vast amount of information to be found on the web? Information management principles and practices, taxonomies, and other controlled vocabularies serve as knowledge management tools that we can use to help organize content and make connections between people and the information they need. This paper focuses on the processes associated with taxonomy development, including how to determine the requirements, how to identify concepts, and how to develop a draft taxonomy. It also covers techniques for validating a taxonomy, processes for incorporating changes within a taxonomy, applying a taxonomy to content, and methods for maintaining a taxonomy over time. Throughout the paper, best practices associated with these process steps are highlighted.*

## **Introduction**

Information overload continues to be a challenge for our end users. For instance, in the corporate world, knowledge workers spend 11 - 13 hours a week searching for and analyzing information. Librarians are faced with solving the information overload problem and we are working to connect end users with the information they need. Simple search is not always the answer because end users tend to enter concepts that are either too broad or are so specific that they do not retrieve key relevant information. What information management strategies can be implemented to improve knowledge sharing? How can librarians better connect information seekers to the vast amount of information waiting to be found?

Information management principles and practices, taxonomies, and other controlled vocabularies all serve as knowledge management tools that librarians can use to help organize content and make connections between people and the information they need. But how do we develop taxonomies and controlled vocabularies? Where do we start? What are the steps involved? What logical processes should we use as we develop taxonomies? We will answer these questions in the body of this paper. In addition, we will provide some best practices associated with these process steps.

## **What is a taxonomy?**

A taxonomy is a controlled vocabulary with each term having hierarchical (broader and narrower) and equivalent (synonymous) relationships. Because of its hierarchical nature, a taxonomy imposes a topical structure on information.

Broader and narrower terms are essential for a browsable hierarchy. The term “aircraft” is broader than terms such as “airplane” or “helicopter”; thus the terms “airplane” and “helicopter” are narrower than “aircraft.” If your terminology is too specific and you cannot retrieve anything, you can move up the hierarchy to less specificity. The reverse is also true; if you are retrieving too much information, seek a more specific term by moving down within the hierarchy. The use of hierarchical relationships is the primary feature that distinguishes a taxonomy from other lesser forms of controlled vocabularies, such as lists and synonym rings.

Equivalent relationships (synonyms) are also embedded in a taxonomy. Synonyms gather together all concepts of a similar nature. The use of the synonym ring helps cast a wide net for information recall.

By using the terms in the taxonomy, you can consistently categorize the information available to you. Using taxonomic subject categories in searches simplifies the search construction process. The searcher does not have to define the subject or master the vocabulary of terms unique to that subject in order to search for information.

## Pre-development considerations of a taxonomy

There are many considerations that need to be taken into account when developing a taxonomy, including facets and intended use, for instance. We are taking for granted that these considerations have already been accounted for and you are ready to create a taxonomy.

Instead, we will concentrate on a seven-step approach to developing a taxonomy. The steps seem rather apparent when displayed in this flowchart, but within each step, there is much work to be done. We hope that as we take the process and break it down step-by-step, the project will not seem so daunting.

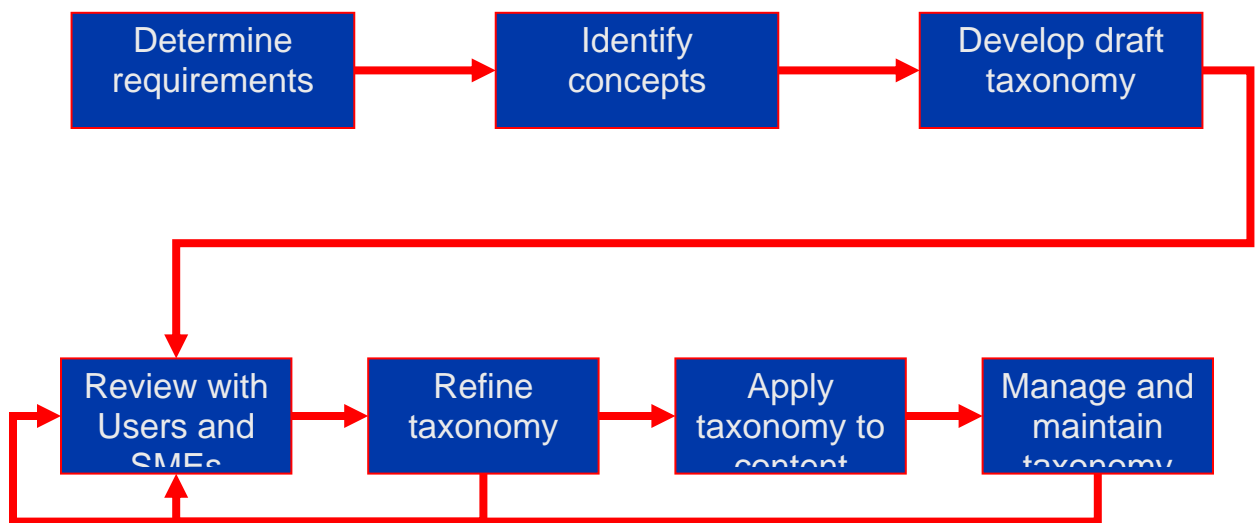


Figure 1. Developing a Taxonomy

Within this seven-step process, there are areas where you will recycle through steps of the process until you are satisfied with the outcome.

### Determine requirements for the taxonomy

Before actually developing the taxonomy, you'll need to define your scope, purpose, and types of content formats. You will also need to identify the target audience and communities who will use it. You might conduct a needs assessment or interviews to identify and focus on content that your users care about.

Some important questions to ask yourself, the team, and/or your end users in order to gain a thorough understanding of your true requirements include these:

What business objective does the taxonomy meet? What problems are the subject matter experts (SMEs) trying to solve and what concepts are important to them? What do the SMEs seem to spend most of their day searching for? Do the SMEs have existing sources

for categorizing information? Are there technology constraints that would have an impact on the taxonomy development?

After determining your requirements, you should document what you've learned in writings such as: a business case document, a scope statement, a project plan, and/or a summary of the users' information needs.

### **Identify concepts within the taxonomy**

The best way to identify concepts for the taxonomy is to use this threefold approach: discover where and what the content is; perform a content inventory; and conduct user interviews.

There are several tasks that need to be completed before moving on. First, determine which collections of content will be included and begin to search for and analyze the content for candidate terms. Look at how the content is structured.

Inventory the content to see what exists and where it is located. How many systems are involved? How broad will the coverage be? How much of a variety exists in the sources? Is there any metadata available that already describes the content?

Analyze the content to determine what the "high value" content is to the users. Through this analysis, you may see some ways to break down the taxonomy into smaller, more easily managed facets. Begin to collect terms that seem to represent concepts that are high value to the organization.

Interview SMEs to understand the problems they are trying to solve and to understand concepts that are important to them.

Learn from existing taxonomic sources such as organizational structures, web pages, and categories used to organize files. Determine if a taxonomy can be purchased that meets the needs of the organization. (For a list of commercially available taxonomies and thesauri, see Taxonomy Warehouse (<http://www.taxonomywarehouse.com/>)). A purchased taxonomy can be modified and serve as a starting point for your own taxonomy. If a commercial taxonomy is not available, and this approach seems appropriate for your subject content, try to gather concepts from external resources, such as societies, associations, and indexes.

If you are organizing a web site, for example, you will want to use search logs to highlight most frequently used search terms; queries with no results, search trends, and most requested items.

Here are some important questions to ask yourself, the team, and/or your end users in order to identify the key concepts you want to include in your taxonomy:

What content do people create? What content is purchased? What repositories does the group manage? Where is the existing content located? Who is responsible for the content? What are the formats of the content? Are the categories concrete like author or geographic location or are they variable and abstract like “risk analyses” and “public relations memoranda”? How is the content structured?

### **Develop a draft taxonomy**

First, develop the upper levels of structure into the major categories. Try not to have more than ten large subject areas; if you have more than that, it will make it difficult to navigate through the hierarchy. One structure might be to organize around major domains (products, human resources, geographies, for instance).

Start broad, not deep, when creating this draft taxonomy. You will want to try to avoid developing for low-level aspects of the business or areas that are outside the interests of the topic.

Once you have the highest-level draft categories (ten or so) for your overarching structure, you can work from the bottom up and from the top down. Top down starts at the general level and focuses on what the collection is about to determine how to organize the top levels of the taxonomy. A bottom up approach helps define how broad or deep the taxonomy needs to be, based on the range of subjects covered by the source materials. A combination of both approaches must be used. If you are working on a team, consider assigning individual categories or sections of the taxonomy to team members.

The terms in the taxonomy should be descriptive enough to be meaningful and unique. For each term chosen, ask “does this term communicate the concept?” When making a decision about which term to use, it is best to select commonly understood terminology as much as possible.

Establish common rules for taxonomy term format, relationships, and structure. This topic could be a paper itself and is thus outside the scope of this paper. Refer to the ANSI/NISO standard, Z39.19 (2005), *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* for authoritative guidelines and conventions. It is an invaluable resource and it will help answer your questions about when to use a singular or a plural noun, for instance. It is best to document the guidelines and standards you will be using so that changes in the future will be made consistently.

### **Review draft taxonomy with users and subject matter experts**

This is one of the iterative steps in the process. By creating a draft taxonomy to give users something to respond to, they will better understand what a taxonomy is and they won't be trying to build it from scratch, which could be a recipe for disaster. Involve stakeholders, subject matter experts, and users from across the business who can

represent end-user interests and who understand the business. Stakeholder and end user agreement is critical to ensure ongoing support so build consensus as much as possible.

Conduct usability studies to help determine if the structure makes sense to the end users. One form of a usability study would be to have several people index the same items. Inconsistencies in indexing can point out problems within the taxonomy. There are other forms of usability studies, again, beyond the scope of this paper.

Some important questions to ask yourself, the team, and/or your end users when reviewing the draft taxonomy include these:

Are the users and subject matter experts able to validate the taxonomy? Does the structure make sense to the users? Does the taxonomy go too deep in any place? Are the major concepts included in the taxonomy? Are there any gaps?

Here are some actions to avoid when developing the draft taxonomy:

- Avoid over thinking. It is very easy to get bogged down in details and wasting time in trying to find the perfect words to use. As long as it is easy to understand for your user group that is what is important.
- Avoid developing unneeded sections. This adds unnecessary bulk to your taxonomy and is not value added to the organization.
- Avoid over-engineering. Early attempts in taxonomy development are often too deep, detailed, or broad. Concepts can be ambiguous, not well understood, or not agreed upon.

## **Refine taxonomy**

Refining a taxonomy is an iterative process. Review user and subject matter expert feedback and incorporate agreed-to changes. Analyze the results of usability studies and incorporate changes as appropriate.

Continue the review and refine cycle to build depth into the taxonomy. How deep the taxonomy should be is dependent upon user feedback. Solicit input from users and subject matter experts during this process and document all changes.

Keep in mind that the taxonomy is a living and growing entity. It is never finished. On the other hand, you must become a good judge of when to stop. Here is a piece of advice: know when to call the taxonomy “good enough” and then move on to the next step! Taxonomies can be very expensive to maintain compared to the value they deliver when they are too detailed.

## **Apply taxonomy to content**

Applying a taxonomy to navigate through web sites helps researchers find materials, customers locate products and services, and knowledge workers locate experts.

Provide guidelines for use, application, and training for users. Taxonomy structure includes the terms and relationships between them and it is important for users to understand this structure and the meaning of the relationships between concepts.

Taxonomies are also applied against documents in file servers or databases. When tagging legacy content, you need to balance the resources required to tag against the value of the content. Determine which content will deliver the most value. There are two approaches for tagging legacy content:

- Selectively tag according to evaluation of the content.
- Automatic tagging of electronic content. An automated classification tool can provide a first cut at tagging electronic content.

Integrate the taxonomy with existing applications such as: search engines in search queries; databases for tagging and searching; filtering programs to personalize alerts and websites; and documents in file servers.

It is possible for a taxonomy structure to work in multiple applications, but most business taxonomies are highly specialized. It is not unusual for an organization to use multiple taxonomies for different functions or applications (human resources, marketing, finance, products, for instance).

Think about what other tools are available for searching for information. How granular you go may depend on whether or not there is a full text search capability available that could be used for very specific terms.

Taxonomy terms are associated with content; so if a search returns content that is inappropriate or in error, analyze for term association. A new term may need to be added to the taxonomy to retrieve the item appropriately. Terms used excessively in indexing or very infrequently are candidates for deletion or modification; because they are ineffective in retrieval.

## **Manage and maintain taxonomy**

From the beginning, establish ownership of the taxonomy. Obtain buy-in regarding who “owns” the taxonomy and who will be responsible for maintaining it? A best practice is to assign a team with the responsibility for ongoing management, maintenance, and further development of the taxonomy.

It is also important to establish governance processes and a change control process. The purpose of the taxonomy team is to develop and maintain the taxonomy, including

handling the change control process (reviewing, approving, and implementing changes). The purpose of a change board is to review and approve major changes to content and structure of the taxonomy. A change board may also review and approve changes to strategy and functionality, and if appropriate, elevate issues to an advisory board for guidance. At the highest level, members of an advisory board serve to provide strategic direction and to promote the taxonomy.

Periodically review content for new concepts and to see if the taxonomy still “fits.” Are there aspects of the business that are not represented in the taxonomy? Because language is dynamic, taxonomies will need to change over time to remain current and of value. New concepts emerge; terminology and usage change; and some terms go out of fashion. Your core business might change, as well. You can capture these changes in language by creating a list of candidate terms. Even if they are not used in future versions of the taxonomy, they might be considered synonyms to another term.

Especially when a taxonomy is integrated into other systems, it is vital to have a process for managing version control. For example, will the taxonomy be updated as new terms are added, almost instantaneously, or will there be a scheduled update cycle that is done every quarter or every six months? Coordinate updates with other information systems using the taxonomy.

It is important to document various aspects of your taxonomy. This may be done on the front end user interface page. For example, you might provide a description that includes the purpose and scope of the taxonomy, the meaning of abbreviations, the use of punctuation and different fonts, etc. You should document rules and authorities that you use for format and relationships and any standards you adhere to. Statements regarding an update policy and last update dates, as well as contact information and any special navigation are also forms of essential documentation.

### **Best practices in taxonomy development**

Understand the difference between what a taxonomy is (a scheme to describe what the content is about (subject)) and what a metadata scheme is (a description of content types, format, rights, etc.).

Plan for the long term; have a vision and strategy for the taxonomy. Determine how searchers will use the taxonomy to navigate and retrieve content or discover information.

Understand your target audience and their requirements. Know the collection of content and the subject areas it covers. Use SMEs whenever possible. Keep your taxonomy simple and efficient with only as much detail as needed. Focus on high value content first.



Keep in mind that taxonomies are not effective at organizing content in new or emerging technologies where understanding and meaning is still developing. Give end users a draft to respond to and incorporate responses into the draft.

Research lessons learned from other companies. Review the literature, talk to colleagues in other companies, and evaluate existing taxonomies to see what works well and what does not.

Build, buy, and re-use existing terminology whenever possible. Look at the terminology your community uses for their file names, main sections of their websites, or search logs. Leverage and reuse your vocabularies and term lists as much as possible, and customize where needed.

Consider that some taxonomies are available for purchase and may meet your needs, such as those available from Factiva or LexisNexis. Adapt publicly available thesauri and standards to meet your requirements. Other controlled vocabularies can be purchased, such as *Medical Subject Headings (MeSH)*, the *NASA Thesaurus*, or *The Getty Thesaurus of Geographic Names*.

Buy pre-built solutions with caution. The majority of organizations find that a commercial taxonomy will not match their unique needs and will need to be modified. They are a good place to start, but often will not meet the needs of scientific, engineering, or government organizations. Any purchased or developed taxonomies you use must be integrated into a consistent taxonomical structure.

Build in flexibility. Make sure your taxonomy is scalable as the volume of content increases. As new content is integrated, the taxonomy needs to be extensible to accommodate any new concepts.

Anticipate major disruptions such as mergers, acquisitions, divestitures, and major changes to business models. These can require changes to the taxonomy structure or addition of major branches. Use metrics, periodic usage review, and document taxonomy design processes to make these disruptions easier to manage. Plan for maintenance resources for ongoing staffing and funding to ensure that your taxonomy does not become a dead-end project.

Understand how tools and technology work together. Beware of vendor hype and learn what the technology truly can and cannot do; what will need to be customized to meet your needs, and what additional costs and time might be associated with customization.

Develop metrics on the value and relevance of the taxonomy to your organization. Analyze usage and try to predict the need for change over time. This will also help with future planning and resources.

Develop strong stakeholder and user relationships. Include subject matter experts in from your industry, business process, and information science, if possible. Include a user-

oriented perspective (user needs and interests) as well as content-oriented perspective. Take into account the way users search for content.

Assign responsibility for ownership and maintenance. Create a partnership between Information Technology (IT) and Information Systems (IS) for support. Plan for ongoing staffing and funding. IT and executives may not know there are experts in their libraries. Leverage your professional know-how and be the leader in your taxonomy development.

## **Conclusion**

By using the seven-step process described above, by asking probing questions and seeking thoughtful answers to them, and by following the best practices of taxonomy development, we hope these steps and advice will save you time and effort as you develop your own taxonomy.