Tagging Strategy

Table of Contents

1   Introduction

This document provides guidelines for content tagging, as well as further taxonomy development and use. It is structured so that it may be used as a reference guide when tagging content. General principles on tagging are provided as well as background on taxonomy facets.

Taxonomies are never finished. They evolve and change in response to the constantly changing needs of the business. Therefore, this document is expected to evolve and change with the taxonomy over time. As terminology is applied by content creators and consumed by end users, the taxonomy and the conventions for using it will require refinement.

1.1   Definitions

This document and others may refer to specific terms related to taxonomy and taxonomy development. For clarity, these terms are defined below.

Term	Definition
Vocabularies & Taxonomies	
Controlled Vocabulary	A managed vocabulary (list of *Terms*) that which are used to tag units of information (document or work) so that they may be more easily grouped together as well as retrieved by a search
Taxonomy	A system for organizing content according to shared characteristics. The shared characteristics are often (but not always) organized into hierarchies.
Facet	An independent branch of an overall Taxonomy, which has its own *Metadata Element* and its own *Controlled Vocabulary* for the *Metadata Values*. Each facet represents a discrete aspect of the content which is of importance to the business (e.g. Products, Market, Topic).
Faceted Taxonomy	A *Taxonomy* structured as a set of independent but related Facets.
Refinement Facet	A refinement facet is used in search and should not be considered equivalent to a taxonomy *Facet*. Search refinement facets are selected metadata fields and values (for example: "Featured" or "New"). Each refinement facet represents a specific perspective on content that is clearly bounded and mutually exclusive.
Navigation Taxonomy	A system for navigating a website's content. The intent of *a Navigation Taxonomy* is to provide access to particular pages and to translate the terms in the taxonomy into user-oriented terms. The terms within a *Navigation Taxonomy* need to relate to the terms of the *Taxonomy* used to tag the content, but do not need to be the same. *Navigation Taxonomy* on a website needs to be very fluid and responsive to users' needs.
Node	A facet can be comprised of one or many taxonomy nodes, or branches, which are groupings of one or many terms that appear in the hierarchy.
Terms	
Term	An entry in a *Controlled Vocabulary* that represents a concept from the taxonomy domain. A *Term* has a Label and may have a unique identifier, a *Scope Note*, and relations to other *Terms*. Also referred to as *Node*. Each taxonomy term has a parent unless it is a *Top-Level Term* and each can have zero or more children.
Top-Level Term	A *Term* that has no parent in a *Taxonomy* hierarchy. A *Top-Level Term* can also be its own node.
Label	The 'name' selected to represent a *Term* in a *Controlled Vocabulary*.
Equivalent Term	All other terms used for a particular preferred term, including synonyms

Term	Definition
	(also known as "nonpreferred terms").
Preferred Term	A subset of all terms that are accepted as the official, enterprise-wide terms to describe items and ideas.
Related Term	Term that is associated with the preferred term, but does not have a direct one-to-one correlation with the preferred term.
Metadata	
Metadata	Data about content. *Metadata* is composed of one or more *Metadata Elements* (aka fields).
Metadata Field (Element)	A basic unit of *Metadata* that describes a particular content property. Each *Metadata Field* (or element) is composed of a name and a value. The name is also known as the field name. When the metadata field is encoded in XML, the name is also known as the element type. Examples are Title, Publication Date, etc.
Metadata Value	The value portion of a Metadata Field, such as "Jane Doe" for an "Author" element, or "2011-05-15" for a "Date" field.
Tagging	
Tagging	The process of assigning Metadata to content, i.e. the process of assigning values from *Controlled Vocabularies* to appropriate *Metadata Fields* defined in a metadata schema. Also called Meta-Tagging.
Scope Note	A definition for a *Term* in a *Controlled Vocabulary*. In addition to defining the *Term* and distinguishing between terms with overlapping meanings in natural language, a *Scope Note* may also provide guidance on *Term* usage, e.g. in which cases it should be used for *Tagging* and when a different *Term* should be used.

1.2   Metadata – Additional Definition

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use and manage a content item or asset. Metadata in the [Client] domain will serve three primary purposes - to help users find information easily and quickly; to co-locate information around specific topics; and to support content management and re-use.

There are three basic types of metadata:

Descriptive metadata, which is usually a type, definition, or title. For example, a specific document stored on a repository might be identified as a case study (type), considered a view on the use of a particular [Client] component in an application (definition), and named "High-Temperature Electronics Pose Design and Reliability Challenges"(title).

Administrative metadata, which is often created automatically by the system. For example, the budget document has a file type (pdf, or more precisely an Adobe Acrobat document), a date of creation, a date of most recent modification, etc.

Structural metadata, which is defined by how the document is organized within a system; that system is usually a hierarchy of child-parent relationships, a collection of multiple hierarchies. For example, a "High-Temperature Electronics Pose Design and Reliability Challenges" document might be part of a larger collection, within "[Client] Dialogue," which contains not only this document, but others which are part of an electronic publication. That is, the document itself may also be seen as an aggregation of smaller documents, such as chapters and appendices.

The primary purpose of all metadata is to enhance the findability of content. For example, descriptive metadata can allow a content item to be found by its title, administrative metadata will allow content

items to be sorted by date, and structural metadata will allow users to collect similar content items in intuitive groups.

This document will primarily focus on the descriptive metadata, which requires direct tagging (as opposed to a title field, which is already present within the system and typically requires no modification).

1.3 General Tagging Principles

This document details specific guidance on tagging and term maintenance (when appropriate) for each metadata field and associated taxonomy facets. However, the principles listed below should govern all tagging activities. Adherence to these principles will help insure consistency and quality of metadata for [Client] content.

- The content described should be faithfully portrayed. Choose terms that represent the content. What is the content "about"? What is the author trying to convey?

- The number of tags used should be limited and focus on quality, not quantity. A general rule of thumb is that, when multiple values are allowed, there should be no more than a maximum of three tags per facet per content piece. Of course, exceptions occur -- particularly for products.

- Tags should be applied consistently across all content.

- Tags should be selected at the most specific level which accurately represents the content. If you find that you are applying the same tags to all your content, then tag to a more detailed level. Tagging systems will allow inheritance of terms if tagging with a hierarchical taxonomy. So, any lower (more narrow) tags will also inherit the terms in the higher (broader) category above. However, this will have some technical implications.

- Tags should not be selected based on minute references to a market, product, etc. Focus on the primary topic(s) of the content.

- Tags should be selected on common usage; terms used should be in accordance with the majority of users.

- Metadata fields that are not required may not apply to specific content items and should be left blank in those cases.

- Do not tag based on a navigation system. Navigation comes and goes, you should not have to retag content whenever your website changes.

1.3.1 Videos & Webcasts

It should be noted that videos will be stored in Brightcove repository, but the tags will live in the CMS. It is planned that archived webcasts will also be stored in Brightcove. Tagging for videos can leverage current (or planned) taxonomy facets as appropriate. In addition, new facets should be created to support video content, including: video language and video type. It is also recommended that a web service (or similar) be create so that when a video is stored in Brightcove, associated CMS tags will flow back to Brightcove. The current assumption is that videos.[Client].com (powered by Brightcove) will continue to exist.

1.3.2 News This Week

Currently, a taxonomy value is used to indicate what content is included as part of a weekly mailing that is sent as part of a My[Client] notification. Going forward, in Sitecore, the following method will be used instead: