

# The 4Cs of Meta Tagging

Implementing a new content management system promises to bring many benefits to an organization's web site and team. However several challenges need to first be overcome before these benefits can be fully realized. Perhaps the least understood of these challenges is the process of meta tagging the site's content.

We're going to assume for the sake of this article that an organization has determined that metadata is a key component to their content management and delivery mechanism and that their website is large, and possibly global in nature. These conditions mean that they are facing a significant tagging effort.

In this article we are going to explore some of the fundamental concepts of metadata and introduce the reader to the primary challenges that a team faces when tagging a large amount of content.

## The Core Issue

Metadata is commonly referred to as data about data. In the context of a web site, it is information about a piece of content that describes and identifies the content in a way that makes it easier to be used, managed and located. Metadata is in fact the life blood of many content management and delivery systems. Without it content can be very difficult to locate in the system and in many cases advanced features like dynamically generated site navigation, related content and spotlight features become impossible to implement.

These systems can be very demanding on the metadata and there is little margin for error. Indeed the meta tagging effort must be executed flawlessly, but this is very hard to do. The crux of the issue lies in the fact that every person's view of the world is shaped by their unique life experiences.

Often, two people will attach entirely different metadata to the same piece of content because of these differing points of view. Let's explore this deeper by imaging a fictitious scenario.



Figure 1

In this scenario, two people, Cindy and Robert, were asked to look at the photograph in Figure 1 and come up with five descriptive words or phrases.

Cindy's results:

1. Brown
2. Vacation
3. Hawaii
4. Palm
5. Coconuts

Robert's results:

1. Tropics
2. Mai Tai
3. Tree
4. Pirates
5. Purple

**Endstate Solutions**  
25075 Blue Ravine Road  
Suite 120-310  
Folsom, CA 95630

888.569.2982  
info@endstate.com

Clearly if this were a piece of content tagged in their content management system, they'd be in trouble. Neither person is really wrong, but the system won't work unless they apply the same metadata to the content item. Let's take a closer look at the results to understand where they come from.

### Extrinsic vs. Intrinsic

Our first step is to better understand the metadata by identifying which type each data item belongs to. Generally speaking, there are two types of metadata, intrinsic and extrinsic. Intrinsic metadata is descriptive information about an item that is generated through an analysis of the item itself. In the example, "tree" and "palm" are clearly items of intrinsic metadata. Extrinsic metadata on the other hand is descriptive information that the content inherits due to its association with something else. Obviously the image does not depict an alcoholic drink, yet "Mai Tai" was chosen as a descriptor because Robert associates palm trees with drinking mai tais.

Now if we group the metadata by type we get:

Extrinsic	Intrinsic
Hawaii	Palm
Vacation	Brown
Coconuts	Tree
Tropics	Purple
Mai Tai	
Pirates	

### The 4Cs

Understanding the two types of metadata gives us important perspective as we analyze the challenges that are inherent in a meta tagging effort. We use a categorization called the "4Cs" to group these challenges.

The first C is called **Completeness**. This is simple; did all of the appropriate metadata get applied to the content item, or was some left off? Cindy and Robert had big problems with completeness. If you look at the results you will see that "Hawaii" appeared in one result set, and "Tropics" in another. Because these pieces of metadata are extrinsic in nature and both are equally subjective, we can't really argue that one is more correct than the other. You could argue though that because they are both correct, they are both needed in order for the tagging to be complete.

Completeness is a common problem when you have multiple meta taggers working together. Unfortunately without complete metadata, a web site will often not function. For example, CNET has a section of their web site dedicated to the review of mobile phones. What would happen if a block on content related to Blackberry™ mobile phones was given to multiple people to meta tag? Some of the people might tag the content with "RIM", the company that makes the Blackberry, while others might tag the content with "Blackberry". None of these people would be incorrect, but they would be incomplete because without both pieces of metadata the content cannot be located by both company and product name. The project's metadata strategy should contain rules to ensure completeness of tagging.

As you may have already determined, completeness can quickly get out of hand. This is where the second C comes into play; **Controlled**. Meta tagging does not begin until a metadata strategy has been defined. One of the primary purposes of a metadata strategy is to provide set of instructions and controls for the tagging team. The foundation of this is a controlled vocabulary. This is essentially an enumerated list of all of the metadata items and values that are allowed. There are many industry standard controlled

**A metadata strategy that uses an ontology or schema and is inclusive of structural, administrative and descriptive metadata should be established prior to a meta tagging effort.**

vocabularies, but often they need to be custom developed for the specific demands of the system.

Had Cindy and Robert used a controlled vocabulary for their tagging, several of the issues they encountered may not have occurred at all. For instance, rather than leaving the meta tagging completely open to their subjective experiences, a controlled vocabulary would have given them a common ground to work from. Also, a controlled vocabulary explicitly limits what is allowed, so “Tropics” might not have been an available value and “Hawaii” may have been the next best match for Robert.

The third C is called **Correctness** and it is most often associated with intrinsic metadata. You may have noticed that Robert included “Purple” as one of his metadata values. This is most likely a piece of intrinsic metadata, but if you look at the picture you won’t really see any purple. Correctness comes into play when a person examines a content item and simply makes a mistake in their tagging. Robert may have been color blind, or perhaps he meant to say “Brown”, but purple was on his mind.

Correctly tagging a piece of content becomes even more challenging when the controlled vocabulary offers multiple, similar options. For example if a controlled vocabulary contains both “Brown” and “Burnt Sienna” or both “Product Brief” and “Product Specification” things can get confusing. Ideally a controlled vocabulary will be constructed to avoid this issue, but at the very least the metadata strategy should give instructions on how to correctly use each metadata item.

The last C in the group is **Consistency**. This comes into play when different people use different metadata and values to describe the same thing. Their intentions are often the same, but the problem lies in the execution. For example, Cindy used the word “Palm” while Robert used the word

“Tree”. It may very well be that Cindy also intended to identify that a tree was the subject of the picture, but the specific type of tree was the first thing that came to mind. Neither person is really incorrect because this is subjective, extrinsic metadata, but they are certainly not consistent with one another.

The issue of consistency is similar to the issue of correctness, especially when the controlled vocabulary offers multiple, similar options. Again, a clearly defined set of instructions will help the people tag the content consistently with one another.

## Conclusion

The concept of the 4Cs is never more relevant than on a meta tagging project that spans multiple geographies. These projects will normally involve teams dispersed across the globe, in different time zones, and with extensive cultural and language differences. Naturally these issues only make it harder to keep things consistent, complete, correct and controlled. But even on smaller projects that are confined to one location, developing a comprehensive strategy to handle these issues is paramount to the success of the meta tagging effort.

Meta tagging is often a critical, but under appreciated aspect of implementing a new content management system. After reading this article you should have a good idea of what to look out for so you can start planning your project.