



The Importance of Controlled Vocabularies



Description of how a controlled vocabulary can enhance a system

[Controlled Vocabularies \(/glossary/29#term11024\)](#) are important to researchers for many reasons:

- Consistency
- Accuracy
- Automation
- Simplification of input
- [Interoperability \(/glossary/29#term11037\)](#)
- Enhancement of searches and [discovery \(/glossary/29#term11027\)](#)
- Completeness
- Long- and short-term management
- Efficient use of time

In many cases, controlled [vocabulary terms \(/glossary/29#term11748\)](#) completely define the allowable content for a particular [metadata \(/glossary/29#term11033\)](#) element.

Also, a controlled vocabulary can be easily incorporated into automated procedures. In a data system, a controlled vocabulary can simplify system input and contribute to quality control by providing users or other systems with a list of allowable entries for the specific [metadata elements \(/glossary/29#term11034\)](#), and can be used to check existing or imported metadata descriptions for consistency and correctness, including spelling and hyphenation.

Controlled Vocabularies as an Interoperability Aid

When a metadata description created by one system can be interpreted by another system, the resource described by the metadata can be used more easily and precisely within both systems.

In spoken language, when we move from one language to another, we need to identify a word in our own language and relate it to a word in another language. We also might need to take a closer look at the word in our language to determine exactly what it means or to define its proper usage. There are times when a word doesn't translate directly into a single word or phrase in another language.

In the context of metadata, a controlled vocabulary is analogous to a language. If the terms in one controlled vocabulary can be translated into the terms used by a second controlled vocabulary, then all metadata descriptions that use the first controlled vocabulary can also be translated to use the second controlled vocabulary. In this way, controlled vocabularies facilitate metadata interoperability.

The different types of controlled vocabularies provide different levels of interoperability. Often when we move from one project to another, we need to identify the metadata descriptions that use one controlled vocabulary and relate these descriptions to another system. We might need to understand more about the terms in the initial controlled vocabulary: what it represents ([glossary \(/glossary/29#term11575\)](#)), how it came to be ([dictionary \(/glossary/29#term11026\)](#)), and what terms are similar ([thesaurus \(/glossary/29#term11055\)](#), [semantic network \(/glossary/29#term11579\)](#), or ontology). There will be times when one term doesn't fit neatly into the second controlled vocabulary. This is where hierarchies and other classifications ([subject headings \(/glossary/29#term11577\)](#), [taxonomies \(/glossary/29#term11578\)](#), and [ontologies \(/glossary/29#term11044\)](#)) become handy.

Example of Controlled Vocabulary Usage

Suppose three different oceanographic research projects are using various vessels or submersibles. In the worst case, we could imagine that none of these projects had a controlled vocabulary. In this case, if someone were to query the data resource to accurately locate all data associated with a particular research vessel like the R/V Moana Wave, they would need to know all the ways "R/V Moana Wave" was represented within the resource, and construct a search query for all of the variations (including the misspelled, misrepresented and nicknamed). This seems nearly impossible!

In a better case, we could suppose each project generated a controlled vocabulary, as shown in the three diagrams below.

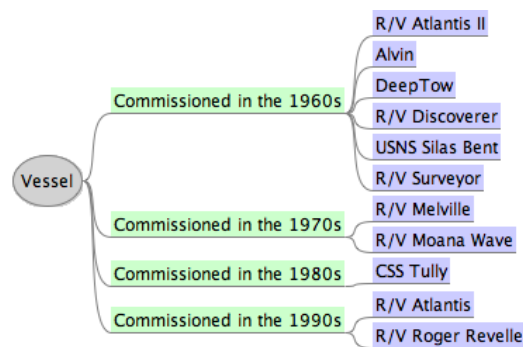
Figure 1



Dictionary

- Each term is articulated with an acronym. (1st entry, blue)
- The acronyms are spelled out in the description. (2nd entry, yellow)
- Additional information about how each term came to be is included in the etymology. (3rd entry, green)

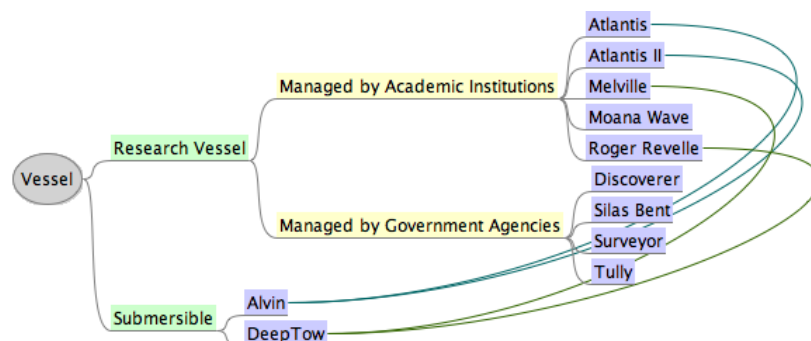
Figure 2



Taxonomy

The actual terms (2nd entry, blue) are placed in a structure, according to the decade in which they were commissioned (1st entry, green).

Figure 3



Ontology

Actual terms (3rd entry, blue) are classified into two major [classes](#) ([/glossary/29#term11023](#)) (1st entry, green), and one subclass (2nd entry, yellow).

Notice the vessels are connected to submersibles, based on the operating institution. This is a complex interrelation that enhances the class hierarchy.

Each of these controlled vocabularies represents the same list of real-world objects (i.e., vessels or submersibles). They are presented as different types of controlled vocabularies, with different terms to represent the real-world objects, and with slightly different accompanying information.

Suppose each project exposed their particular controlled vocabulary to a search engine and that translations existed between the vocabularies. The search engine may provide a drop-down menu of platform names to expedite the user searches. When a user needs to identify all data associated with the R/V Moana Wave, they could use a drop-down menu to select that particular ship.

The example above also illustrates the [value](#) ([/glossary/29#term11041](#)) of adopting established controlled vocabularies, instead of developing a local vocabulary. Each of these three controlled vocabularies is a representation of the same set of

real-world objects, but three different projects took the time to develop a unique controlled vocabulary.

One or more of the locally developed controlled vocabularies might not be exhaustive, and not all three contain the same information. If the three programs collaborated and developed a single controlled vocabulary, this authoritative controlled vocabulary could be managed centrally. The controlled vocabulary would be more complete, and thus would be much stronger, possibly with less effort by any individual program.

Posted November 7th, 2008

SUGGESTED CITATION

[Bermudez, L. \(/taxonomy/term/11564\)](#), [Montgomery, E. \(/taxonomy/term/11565\)](#), [Miller, S.P. \(/taxonomy/term/11566\)](#), [Neiswender, C. \(/taxonomy/term/11567\)](#), [Isenor, A. \(/taxonomy/term/11568\)](#) 2011. "The Importance of Controlled Vocabularies." In *The MMI Guides: Navigating the World of Marine Metadata*. <http://marinemetadata.org/guides/vocabs/vocimportance>. Accessed October 21, 2014.

[Have suggestions? Please give us your Feedback \[+ \]](#)